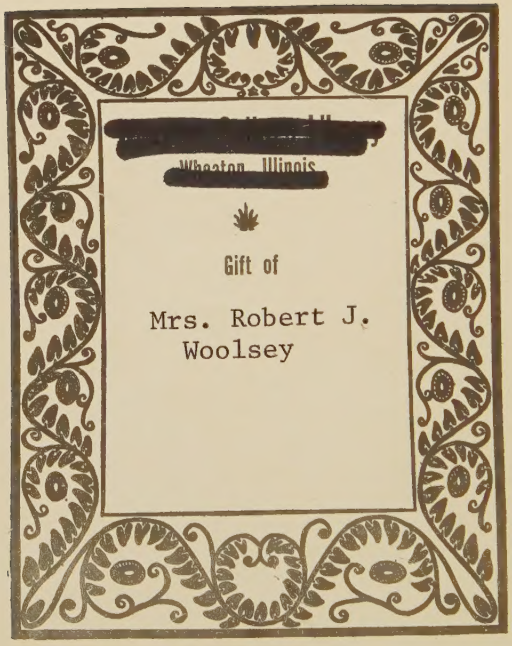

[redacted]
Wheaton Illinois
Gift of
Mrs. Robert J.
Woolsey

134r750
RJW

Probability

and its Engineering Uses

BOOKS FROM

BELL TELEPHONE LABORATORIES, INC.

THEORY OF VIBRATING SYSTEMS AND SOUND. *By* IRVING B. CRANDALL, *Late Member of the Technical Staff, Bell Telephone Laboratories.*

CONTEMPORARY PHYSICS. *By* KARL K. DARROW, *Member of the Technical Staff, Bell Telephone Laboratories.* Second Edition.

SPEECH AND HEARING. *By* HARVEY FLETCHER, *Acoustical Research Director, Bell Telephone Laboratories.* With an Introduction by H. D. ARNOLD, *Director of Research, Bell Telephone Laboratories.*

PROBABILITY AND ITS ENGINEERING USES. *By* THORNTON C. FRY, *Member of the Technical Staff, Bell Telephone Laboratories.*

ELEMENTARY DIFFERENTIAL EQUATIONS. *By* THORNTON C. FRY. Second Edition.

TRANSMISSION CIRCUITS FOR TELEPHONIC COMMUNICATION. METHODS OF ANALYSIS AND DESIGN. *By* K. S. JOHNSON, *Member of the Technical Staff, Bell Telephone Laboratories.*

A FUGUE IN CYCLES AND BELS. *By* JOHN MILLS, *Member of the Technical Staff, Bell Telephone Laboratories.*

TRANSMISSION NETWORKS AND WAVE FILTERS. *By* T. E. SHEA, *Special Products Engineer, Bell Telephone Laboratories.*

RHOMBIC ANTENNA DESIGN *By* A. E. HARPER.

THE APPLICATION OF ELECTROMECHANICAL IMPEDANCE ELEMENTS IN THE TRANSDUCERS AND WAVE FILTERS. *By* W. P. MASON.

PUBLISHED BY

D. VAN NOSTRAND COMPANY, INC.

519.2
F947p

Probability

and its Engineering Uses

By

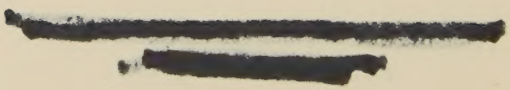
THORNTON C. FRY, PH.D.
Member of the Technical Staff
BELL TELEPHONE LABORATORIES, INC.

THIRD PRINTING

NEW YORK
D. VAN NOSTRAND COMPANY, INC.
250 FOURTH AVENUE

COPYRIGHT 1928,
BY
D. VAN NOSTRAND COMPANY, INC.

All rights reserved, including that of translation into the Scandinavian and other foreign languages



Printed in U. S. A.

"Coincidences, in general, are great stumbling-blocks in the way of that class of thinkers who have been educated to know nothing of the theory of probabilities: that theory to which the most glorious objects of human research are indebted for the most glorious of illustration."

—The Murders in the Rue Morgue.

Digitized by the Internet Archive
in 2025

https://archive.org/details/bwb_W9-DJN-969

ACKNOWLEDGMENTS

THIS textbook is the outgrowth of a set of notes originally prepared for one of the "Out-of-Hour Courses" of Bell Telephone Laboratories, and subsequently revised for use in a course of lectures delivered at the Massachusetts Institute of Technology in its Department of Electrical Engineering. I am deeply indebted to both these institutions for arrangements whereby it was possible to try out the material under actual classroom conditions — a rare opportunity for a mathematician engaged in industrial work.

I am also indebted in various ways to a number of individuals:

To Dr. W. A. Shewhart and Mr. L. A. MacColl, of Bell Telephone Laboratories; to Mr. E. C. Molina and Dr. H. Nyquist, of the American Telephone and Telegraph Company; and to Mr. Arne Fisher, of the Western Union Telegraph Company, for valuable advice and criticism.

To Mr. R. A. Fisher, of the Rothamsted Experimental Station, and to his publishers, Messrs. Oliver and Boyd, of Edinburgh, for permission to use the tables contained in Appendices VIII and IX, which are copied, except for slight changes in notation, from his excellent monograph on *Statistical Methods for Research Workers*.

To Miss Clara L. Froelich, of Bell Telephone Laboratories, for assistance without which the book could not well have been written. Not only are the figures almost entirely the work of her hands, but she has assisted materially in preparing the copy for the printer, among other things checking the entire manuscript for mathematical errors. It cannot be hoped that no errors remain, but her efforts have greatly reduced their number.

THORNTON C. FRY.

NEW YORK,
February 15, 1928.

CONTENTS

CHAPTER I

Introduction

SECT. PAGE

1. The Fundamental Subject Matter of Probability 1
2. How Probability is Measured; The Unit of Measure 3
3. How Probability is Measured; The Fundamental Axioms and Conventions 4
4. How Probability is Measured; The Measure of Probability Defined . . 7
5. How Probability is Measured; Shortcomings of the Definition 8
6. Final Remarks 9

CHAPTER II

Permutations and Combinations

7. General Laws of Composition of Events 12
8. Definitions 14
9. Application of the General Laws of Composition to Permutations and Combinations; Some Typical Examples 16
10. Application of the General Laws of Composition to Permutations; General Theorems 18
11. Factorials; The Gamma Function 20
12. Restatement of General Theorems in Permutations 25
13. Application of the General Laws to Combinations 25
14. Some General Properties of C_n^m; Pascal's Triangle 27
15. Some General Properties of C_n^m; The Binomial Theorem 29
16. The Solution of More Complicated Problems 32
17. A Complicated Problem in Permutations 33

CHAPTER III

Elementary Principles of the Theory of Probability

18. Complementary Probabilities 39
19. Unconditional Probabilities 39
20. Conditional Probabilities 43
21. Compound Probabilities 48

SECT. PAGE
22. Alternative Compound Probabilities 53
23. Some Instructive Illustrations; The Psychic Research Problem . . . 57
24. Some Instructive Illustrations; A Generalization of the Psychic Research Problem 58
25. Some Instructive Illustrations; The Problem of Independent Trials . . 62
26. Some Instructive Illustrations; A Generalization of the Problem of Independent Trials 64
27. Some Instructive Illustrations; A Typical Urn Problem 65
28. Some Instructive Illustrations; Another Typical Urn Problem 69
29. Some Instructive Illustrations; A Problem in Matching 70
30. Some Instructive Illustrations; An Example in Computation 73
31. Some Instructive Illustrations; Another Urn Problem 77
32. Some Instructive Illustrations; Another Urn Problem 78

CHAPTER IV

Probability and Experiment; Bernoulli's Theorem

33. Introductory Remarks 82
34. Limits and Things which Approach Them 83
35. The Upper Bound of a Set 86
36. Regarding Probability as a Limit 88
37. Regarding Repeated Independent Trials 91
38. The Limiting Condition as the Number of Trials is Greatly Increased . . 93
39. Bernoulli's Theorem 97
40. Résumé 101
41. Mathematical Justification 102
42. Stirling's Formula 103
43. Another Approximate Formula 107
44. Justification of the First Half of Bernoulli's Theorem 108
45. Justification of the Second Half of Bernoulli's Theorem 110
46. Regarding the Experimental Determination of Probability 112
47. The Multiplication Theorem 113

CHAPTER V

Probability and Experiment; Bayes' Theorem

48. The "Life on Mars" Paradox 117
49. Bayes' Theorem 119
50. Some Instructive Illustrations; Bertrand's "Box Paradox" 121
51. Some Instructive Illustrations; Another Urn Problem 123
52. Some Instructive Illustrations; The Bad Penny 125
53. The Uses to which Bayes' Theorem may be Put 127
54. Some Instructive Illustrations; An Elementary Problem in Sampling . . 129

CHAPTER VI

Distribution Functions and Continuous Variables

SECT. PAGE
55. The Random Choice of a Point on a Line Segment 133
56. A Paradox Associated with the Random Choice of a Point on a Line Segment 138
57. Extension of the Significance of the Preceding Paragraphs 140
58. Distribution Functions for Continuous Variables 141
59. A Variable Which is Not Distributed at Random 143
60. Distribution Functions Derived Empirically 144
61. Distribution Functions in Many Variables 147
62. Some Examples of Change of Variable in Distribution Functions . . . 150
63. Change of Variable in Distribution Functions 153
64. Derivation of a Distribution Function for the Velocities of Gas Molecules 163
65. Some Instructive Illustrations; Change of Variable in Maxwell's Equation 165
66. Information that can be Derived from (73) and (75) 167
67. Some Instructive Illustrations; A More Complicated Jacobian 171
68. The General Significance of the Jacobian 173

CHAPTER VII

Averages

69. Definition of an Average 177
70. Mathematical Expectation 177
71. Derived Averages and Expectations 182
72. Average Values of Continuous Variables 183
73. The Median. 187
74. The Deviation 188
75. Résumé 190
76. Some Instructive Illustrations; The General Case of Independent Trials . 192
77. Some Instructive Illustrations; The General Case of Dependent Trials of the Sort Discussed in § 27 193
78. Some Instructive Illustrations; A Dice Problem 194
79. Some Instructive Illustrations; The St. Petersburg Paradox 194
80. The Expectation of a Probability 199

CHAPTER VIII

The Distribution Functions Most Frequently Used in Engineering

81. Introductory 205
82. Distribution Functions for Discrete Variables; The Binomial Law and Various Approximations to It 206
83. Distribution Functions for Discrete Variables; The Poisson Law as a Limiting Case of the Binomial Law 214

SECT. PAGE
84. Definitions of the Phrases "Individually at Random" and "Collectively at Random" 216
85. Second Demonstration of the Poisson Law 220
86. Discussion of the Poisson Law; Problems to Which It is an Appropriate Solution 227
87. Discussion of the Poisson Law; Variable Traffic Density in a Telephone Exchange 232
88. Discussion of the Poisson Law; The General Problem of β-Ray Emission 235
89. An Approximation to the Poisson Law 237
90. The Normal Law 241
91. Empirical Families of Curves; Pearson's Curves 244
92. Empirical Families of Curves; The Gram–Charlier Series 251
93. Gram–Charlier Approximations to the Binomial and Poisson Laws . . . 255
94. Empirical Families of Curves; Transformation of Variable 261

CHAPTER IX

Curve Fitting

95. The Primary Problem 265
96. The Hypothetical "Universe of Data," or "Population," and the "Sample" 266
97. The Accepted Criterion as to Goodness of Fit 268
98. Some Instructive Illustrations; The Biased Die 271
99. Discussion of Example 49 275
100. Some Instructive Illustrations; Weldon's Dice Data 277
101. An Approximation to the General Multinomial Law 280
102. The Measure of the Goodness of Fit, $P(>\chi^2)$ 285
103. The Solution of Examples 50 and 51 289
104. Résumé of the Test of Goodness of Fit 291
105. Some Instructive Illustrations; Some Telephone Data 294
106. Some Instructive Illustrations; Chips Drawn from a Normal Universe . . 296
107. The Determination of a Suitable Distribution Function when no Theoretical Formula is Known 297
108. Some Instructive Illustrations; A Reconsideration of Weldon's Dice Data 302
109. Sheppard's Corrections to Moments Computed from Classified Data . . 310
110. The Distribution of Statistics 312
111. Control Charts 315

CHAPTER X

The Theory of Probability as Applied to Problems of Congestion

112. Introductory 321
113. Notation 324
114. General Assumptions 324

SECT. PAGE
115. Some Problems of Lost Traffic 326
116. The Elementary Probabilities; Lost Calls Held 329
117. Introduction of the Assumption of Statistical Equilibrium; Lost Calls Held 331
118. The Probability Formulæ Corresponding to Assumptions 7 and 10 . . . 334
119. The Probability Formulæ Corresponding to Assumptions 8 and 10 . . . 336
120. The Probability Formulæ Corresponding to Assumptions 9 and 10 . . . 338
121. The Elementary Probabilities; Lost Calls Cleared 339
122. The Probability Formulæ Corresponding to Assumptions 7 and 11 . . . 340
123. The Probability Formulæ Corresponding to Assumptions 8 and 11 . . . 342
124. The Probability Formulæ Corresponding to Assumptions 9 and 11 . . . 343
125. Recapitulation of Formulæ 344
126. Numerical Comparison of Formulæ; The Dependence of the Probability of Loss upon the Number of Sources when the Traffic Density of the Group is Held Constant 346
127. Numerical Comparison of Formulæ; The Dependence of the Allowable Traffic Density upon the Number of Sources when the Proportion of Loss is Fixed 349
128. Charts for Purposes of Computation 351
129. Some Hunting Problems 356
130. Individual Hunting from a Normal Position 358
131. Individual Hunting with Stay-Put Switches 363
132. Group Hunting with Stay-Put Switches 364
133. The Problem of Double Connections 369
134. Delays in Awaiting Service. 372
135. Calls of Equal Length at Non-Cooperative Channels; The Probability of Congestion j 374
136. Calls of Equal Length at Non-Cooperative Channels; The Expected Delay 376
137. Exponential Distribution of Holding Times; Delays at Cooperative Groups of Channels 378
138. Exponential Distribution of Holding Times; The Probability of Congestion j 380
139. Exponential Distribution of Holding Times; The Expected Delay . . 382
140. Exponential Distribution of Holding Times; The Probability of a Delay Exceeding the Length τ if Calls are Served in the Order in which They Originate 383
141. Exponential Distribution of Holding Times; The Proportion of Delayed Calls 385
142. Exponential Distribution of Holding Times; The Expected Delay of Delayed Calls 386

CHAPTER XI

Fluctuation Phenomena in Physics

143. Introductory Remarks; Notation 389
144. The Dynamics of Collision 390
145. The Probable Flux Across a Surface 393
146. Change of Velocity Class 395

SECT. PAGE
147. The Fundamental Equation of the Kinetic Theory of Gases 399
148. The H-Function 400
149. Maxwell's Law of Velocities 404
150. Pressure 406
151. The Expectation of the Distance Travelled by a Molecule 408
152. Number of Collisions 409
153. Expected (" Mean ") Free Path 409
154. Density Fluctuations 410
155. The Rapidity of Density Fluctuations 414
156. The Schottky Effect 417

APPENDICES

I. The Factorials of Integers 427
II. The Logarithms of Factorials 429
III. The Binomial Coefficients C_n^m 439
IV. The Normal Error Function 453
V. The Normal Law, Its Integral, and Its Derivatives up to the Sixth . . 456
VI. The Poisson Formula, ${}^{\infty}P'(j)$ 458
VII. The Poisson Formula, ${}^{\infty}\Pi'_{\nu}$ 463
VIII. Pearson's Criterion of Goodness of Fit, $P(>\chi^2)$ 468
IX. Standard Deviations of Important Statistics 470
X. Criteria for Choice of Distribution Curves 470
XI. Statistics of the More Common Distribution Laws 471

INDEX 473

CHAPTER I

Introduction

§ 1. *The Fundamental Subject Matter of Probability*

It is the fundamental purpose of the theory of probability to answer such questions as: What is the probability of tossing an ace with a die? What is the probability that Christmas falls on Monday? What is the probability that the next child born in New York is a girl? What is the probability that Friday falls on Sunday? What is the probability that twenty sheets of paper in a package of 500 differ from the average by more than 1 per cent in thickness?

The subject deals with other questions—about "expectation," "correlation" and the like—but they are all subordinate to the question, What is the probability of a certain phenomenon? Whatever the subject matter, the phenomenon of which the probability is sought is called an "event."

Asking for the probability of an event in itself implies some degree of doubt as to its occurrence; that is, it implies the possibility that the event may not occur. Of course, there are certain causative or controlling factors which determine whether or not the event will occur. Divine intervention is not anticipated; and with sufficient information the answer to the question would be either "It is certain to occur" or "It is certain not to occur."

Thus, in the case of Friday falling on Sunday, the answer is "It is certain not to occur," for *we know* that the thing cannot occur. Moreover, if the question, What is the probability that Christmas falls on Monday? were asked about Christmas of this year, it would be possible to look it up in a calendar and find out on what day it actually falls. As it

either does or does not fall on Monday, the answer would then be a statement of fact, not of likelihood. *Unless a certain amount of ignorance exists such questions are trivial.*

But asking for the probability of an event implies *more* than mere ignorance. It also implies that ignorance is sometimes of less consequence than other times. Take, for example, the questions, What is the probability that the next child born in New York is a girl? and What is the chance that the next ten children born in New York are all girls? The *fact* is not known in either case; but no doubt exists that ignorance is more serious in the first case than in the second. One event is less in doubt than the other. From this point of view the probability of an event evaluates the importance of our state of ignorance regarding it.

This illustration reveals two phases of the intuitive concept of probability. One is that either event *may* occur: that is, the next child *may* be a girl, or the next ten children *may* be girls. This phase is purely *qualitative*. The other is that the first event is decidedly more probable than the second. This phase is *quantitative*: some probabilities exceed others.

Consider also the matter of Christmas falling on Monday. There are seven days in the week and it is a matter of common knowledge that there is nothing in the arrangement of the calendar which tends to favor one of these days rather than another. This thought finds expression in the phrase: Christmas is "just as likely" to fall on Monday as on any of the other days. Moreover, we find it natural to say that Christmas is twice as likely to fall on either Monday or Tuesday as it is on Monday alone; and that it is three times as likely to fall on Sunday, Monday, or Tuesday, as on Monday alone.

This illustration reveals two more intuitive ideas associated with the concept of probability. One is the idea of "equally likely." The other is the idea that, under certain circumstances at least, the probability of *one or the other* of several events is the sum of their separate probabilities.

§ 2. *How Probability is Measured; The Unit of Measure*

These ideas are of a purely intuitive nature. They are merely an appraisal of that common understanding of the word "probability" which makes it an element of speech. Define it we cannot, any more than we can define "length" or "time" or "value" or other quantitative concepts; but we *can* define a method of measuring it, just as in the case of "length" or "time" or "value." And just as we are accustomed in speaking of "length" to substitute the number for the fact, so, too, we shall generally, after we have passed on to the mathematical phases of our discussion, use the word "probability" for what we should, if we were exact, speak of as the "measure of probability." For the present, however, we maintain the distinction, and, admitting our inability to define probability, seek for a method of measuring it.

Such a method flows naturally from the ideas already presented. Using again the Christmas illustration, the numerical measure of the probability of Christmas falling on Monday may be denoted by p. The value of p is unknown, but certain relations into which it enters have already been stated. For instance, the chance of Christmas falling on Tuesday is also p, and the chance of it falling on any other day is the same. Moreover, we have said that it appears natural to say that the chance of it falling on one or the other of several days is the sum of the probabilities for the separate days; the probability of it falling on one or the other of the seven days of the week is therefore $7p$. But it is certain that Christmas falls on some day of the week: therefore $7p$ must be the number which represents certainty.

What number shall be chosen for this purpose is purely optional, although some numbers may seem more appropriate than others. For example, infinity might seem peculiarly suitable, because it seems natural to say that the event is "infinitely likely to occur." But if infinity is chosen, the equation

$$7p = \infty$$

is obtained; and this requires that p also be infinite. Thus the choice leads to the logical absurdity that the chance of Christmas falling on Monday is represented by the same symbol as certainty, though it does not accord with the idea of certainty.

Unity is the only other number which recommends itself to denote certainty. It leads to the equation

$$7p = 1,$$

from which it is found that $p = \frac{1}{7}$. This value does not violate intuitive ideas, and is therefore more satisfactory than the other. Thus it has become customary to adopt unity as a symbol for certainty. As a consequence of this choice all probabilities are confined to the range of proper fractions, including the end points 0 and 1 which represent *impossibility* [1] and *certainty*,[1] respectively.

§ 3. *How Probability is Measured; The Fundamental Axioms and Conventions*

The above illustration contains, by implication at least, the essential ideas needed for a general definition of the measure of probability. But before proceeding to such a general definition, it is desirable to sort out and restate, as best we can, the intuitive ideas (or axioms) upon which the illustration was based. They are:

AXIOM I.—*The question, What is the probability that the event A occurs, has an answer.*

AXIOM II.—*This answer is quantitative; that is, it can be stated in terms of a unit of measure and a ratio (a pure number).*

AXIOM III.—*If two events differ in no other known pertinent attribute than identity, they are equally likely.*

CONVENTION I.—*The unit of measure is certainty.*

[1] In § 56, certain remarks are made regarding the use of these words.

CONVENTION II.—*The scale of measure is to be so chosen that the probability that either A or B happens is the sum of their separate probabilities, so long as the events A and B are mutually exclusive; that is, so long as it is impossible for both of them to happen.*

The third of these axioms requires some discussion. Since the conception of "equally likely" events is intuitive, it cannot be defined, just as other intuitive concepts such as Time, or Sweet, or I, cannot be defined. It is possible, however, by intelligent consideration to give them a greater depth of meaning. Put it this way. Defining an expression enables one to learn *what it means*. We cannot do this with intuitive ideas; but intelligent discussion may enable us to appreciate more fully *what we mean by them*. Thus Axiom III can in no way be called a definition of "equally likely," but it is consistent with the idea which that expression conveys and may even be an aid in checking doubtful cases.

If the Christmas illustration is viewed in the light of this statement, there are seven possible events: Christmas may fall on any one of the seven days of the week. These events differ in *identity*. Otherwise they could not be thought of as distinct events at all. The days themselves differ in other known attributes: Sunday is a day when people go to church, Monday is wash-day, Election Day falls on Tuesday, and Saturday is (or once was) pay-day. Of necessity the events themselves partake to some extent of these attributes; for instance, "Christmas falls on Monday" partakes of the attribute of the day and may also be phrased "Christmas falls on wash-day." These attributes, however, are not *pertinent* to the question at hand: our state of ignorance would be just as important—and no more so—if they were unknown, or even untrue. If habits changed and Thursday become the conventional wash-day, Christmas would still be *just as likely* to fall on Monday as before *and no more so*.

There may be other attributes to these days which are pertinent, but which are unknown. For instance, if the

question is directed at Christmas of this year, one of the days of the week possesses the attribute of being " the day of the week on which Christmas does fall." This is obviously an essential attribute; but so long as it is unknown the probability that Christmas falls on Monday is unaffected.

It may be argued that Axiom III proceeds in a circle because " pertinent attributes " means merely those which influence the likelihood of the event. This is true, and it would be a valid objection to a *definition*; but it must be remembered that III is *not* a definition: it is merely an attempt to state in other words the intuitive meaning of the phrase " equally likely."

And finally, a word of caution should be said about confusing the absence of any pertinent difference between two events, with lack of sufficient knowledge to evaluate the importance of the difference. I ask, If I receive just one telegram to-day, what is the chance that it will be between 1 and 2 o'clock in the afternoon? There are obviously 24 hours in the day, and I do not know what the probability is for any one of them. Shall I therefore assert that they are equally likely? Obviously not; there are vastly more people awake during the hour in question than between 1 and 2 in the morning, for example, and this certainly is a pertinent difference so far as the likelihood of a telegram being received is concerned. In a later section (§ 48) we shall again refer to this point, which is the fundamental error in a widely quoted paradox.[1]

[1] There are two schools of thought, calling themselves "insufficient reasonists" and "cogent reasonists," both of whom accept Probability as an *a priori* fact, but whose ways part on the "definition" of the term "equally likely." The insufficient reasonists say, as we have done (though we do not call it a definition), that two events are equally likely if there is no reason to think them otherwise. The cogent reasonists say they are equally likely if there is a cogent reason for thinking them so.

In so far as one who regards the concept of "equally likely" as intuitive can be said to belong to either school, I am an insufficient reasonist: to me the most " cogent " reason for thinking two things equally likely is the absence of any reason for thinking them otherwise. Therefore the paradox in question, which in one form or another is always used to phrase the objection of the cogent reasonist to the other point of view, is naturally an object of concern.

The present would be the proper time to consider it, except that the paradox itself makes use of certain facts which we shall not be able to regard as established until the end of Chapter IV. We must therefore postpone its consideration until that time.

Before closing this section a word should also be said about the conventions by means of which the scale of measure is defined. Regarding this scale it has already been agreed that unity shall represent certainty, and zero impossibility. The end-points of the scale are therefore fixed. The method of division is also provided for by Convention II; but it must be emphasized that this convention is limited to *mutually exclusive* events; that is, to events of which one at most can happen. Why such a limitation is necessary will become evident from a simple example. If all days but Sunday are "week-days," the probability of Christmas falling on a week-day is $\frac{6}{7}$, for it is the sum of the probability of Christmas falling on Monday, on Tuesday, and so on. But the probability of it falling "either on Monday or on a week-day" is not therefore $\frac{6}{7} + \frac{1}{7} = 1$. Such a result is absurd.

The difficulty is not peculiar to probability theory, but is a fundamental one met in all methods of measurement by direct comparison. We may say, "A rod is two units long if it contains two parts each a unit in length," but in so saying we obviously mean mutually exclusive parts. We can cut off a unit of length from either end of a bar which is only 1.1 units in length, both of which are therefore contained in it; but we do not therefore conclude that its length is 2.

§ 4. *How Probability is Measured; The Measure of Probability Defined*

We are now prepared to make use of these axioms and conventions in the formulation of an exact definition of the "measure of probability." We begin by noting that the argument of § 2, from which the number $\frac{1}{7}$ was derived, made use of the following facts:

(1) The event for which the probability was sought (Christmas falling on Monday) is one of a group of events (corresponding to the various days of the week).

(2) The events are mutually exclusive.

(3) They are equally likely.

(4) The group is "complete"; that is, one or the other of the events *must* happen.

These four facts made it possible to set up the equation $7p = 1$, from which the probability of the event happening was obtained in the form $p = \frac{1}{7}$.

If the complete group had contained m events, the probability of a particular one occurring would obviously have been obtained from the equation $mp = 1$, and would therefore have been $p = 1/m$.

To the question, What is the probability that Christmas falls either on Sunday, Monday, or Tuesday? Convention II may be applied. The answer is the sum of the probabilities of the separate events contained in this group of three. As the probability of each of these is $\frac{1}{7}$, the answer must be $\frac{3}{7}$. More generally, if there were m members in the complete group and it were required to find the probability that some one of a smaller group of n events took place, it would only be necessary to add together n fractions each of the value $1/m$. The answer is therefore n/m. This leads at once to the definition:

If a subgroup of n events is contained in a complete group of m mutually exclusive and equally likely events, the probability of some one of the subgroup occurring is measured by n/m.[1]

§ 5. *How Probability is Measured; Shortcomings of the Definition*

The definition of the measure of probability is not without its shortcomings. In the first place, the definition virtually

[1] This definition is often stated in somewhat different terms: Each event in the complete group is called a "case," the events in the desired subgroup being said to be "favorable," and the subgroup itself being called "the event," regardless of whether it is an individual event or some combination of events. Using the words in this sense, the definition appears in the form:

The probability of an event happening is the ratio of the number of favorable cases to the total number of cases, all cases being equally likely.

We should add that, although we have conformed to the usual custom of calling this statement a "definition," it is in fact a theorem.

implies that both n and m are finite; otherwise "indeterminate forms" arise. This difficulty is a superficial one, however, for infinite quantities never have a meaning except as the result of a limiting process, and the same limiting processes are applicable to the ratio n/m as to any other number.

In the second place, a similar difficulty arises when we ask for the probability that a shot will miss its mark by a distance of between 11 and 12 feet; for distance is a continuous variable, which raises difficulties similar to those met in defining irrational numbers. We shall find, however, when the occasion arises, that we are capable of overcoming this difficulty also.

There is, however, another difficulty of a much more fundamental sort. If we ask the question, What is the probability that the next child born in New York is a girl? it is impossible to build up any group of events which satisfies the conditions of our definition, for though the group "boy, girl" is complete and though its events are mutually exclusive, they are known *not* to be equally likely. There are many examples of this sort: almost everything about which "statistics" are taken would be suitable as an illustration. In fact, among the questions which arise in many fields of research there are so many more to which the definition cannot be applied than there are to which it can, that many statisticians have been led to seek a foundation for the whole subject of probability in the gathering of statistics, rather than in pure *a priori* logic. While I am convinced that their definition is untenable as a foundation upon which to build a logical structure, we shall find that it —or at least something very like it—will serve an excellent practical purpose in overcoming the shortcomings of our definition.

§ 6. *Final Remarks*

Once more the importance of the words "complete," "equally likely" and "mutually exclusive" in the definition, and of "mutually exclusive" in Convention II, must be stressed; and to make more emphatic the fact that human

nature is prone to forget them, a mistake of a great mathematician will be used to point the remarks.

D'Alembert, when asked for the probability that "heads" will appear at least once in two throws of a penny, argued as follows: Heads appear either first, last, or not at all. There are thus three events, two of which are included in the subgroup desired. Hence the chance of heads appearing during the two throws is $\frac{2}{3}$.

But if "heads first" (or last) means "heads then and tails the other time" the group is not complete, for "heads both times" is also possible; while if "heads first" (or last) means "heads then, no matter what appears the other time" the events are neither equally likely nor mutually exclusive: not equally likely because they differ in the *essential* attribute that two depend on the result of a single throw only, while the other combines the results of two throws; and not mutually exclusive because "heads both times" is included both in "heads first" and in "heads last." Actually, as d'Alembert thought of the problem, the question was answered if heads appeared on the first throw, and a second throw was not needed; hence to him "heads first" meant "heads first no matter what happens last"; while "heads last" meant "tails first and heads last." His group, then, was complete and mutually exclusive; but the events were not equally likely.

Fortunately a suitable group can be found. It is

heads—heads,
heads—tails,
tails —heads,
tails —tails.

As three of these events produce at least one head, the correct answer is $\frac{3}{4}$.

* * *

The above discussion is far from complete, and raises a number of questions of a logical nature, the attempt to answer which would be of interest if it were consistent with the main purpose of the text. These questions, however, must be

passed by, and we take up instead a review of certain algebraic laws that will frequently be needed.

References for Outside Reading

1. Whitworth: *Choice and Chance*, pp. 118–130.
2. Chrystal: *Algebra*, Vol. II, pp. 566–580.
3. Coolidge: *Probability*, Chapter I.
4. Venn: *Logic of Chance*, Chapters I and IV.
5. Keynes: *A Treatise on Probability*, Chapters I–VII.
6. Laplace: *Théorie Analytique des Probabilités*, Introduction, as far as heading " De l'Espérance " (pp. i–xiii in the edition of 1814).
7. Czuber: *Wahrscheinlichkeitsrechnung*, Vol. I, pp. 1–21.

CHAPTER II

Permutations and Combinations

§ 7. *General Laws of Composition of Events*

The study of permutations and combinations rests upon two general laws regarding the composition of events.

First Law of Composition.—*If an event A can happen in m ways and an event B can happen in n other ways, "either A or B" can happen in m + n ways.*

This law is so simple that it scarcely requires proof. A few illustrations will serve to establish its validity. Suppose there are three ways of going from New York to Philadelphia and two ways of going from New York to Boston. Then the number of ways of going *either* to Philadelphia *or* to Boston is obviously 3 + 2 or 5. This is in agreement with the law, but it does not indicate why the ways in which the event *B* happens must be different from the ways in which *A* happens, as is implied by the word "other." A second illustration will make this point clear.

Suppose there are three routes to Philadelphia, one of which leads through Princeton, and that there is no *other* route to Princeton. Then although there are three routes to Philadelphia and one to Princeton the number of ways of going "either to Philadelphia or to Princeton" is not four, but three. It is evident that the word "other" is an essential part of the law.

Second Law of Composition.—*If an event A can happen in m ways and thereafter an event B can happen in n ways, "both A and B" can happen in this order in mn ways.*

If a penny is tossed it may fall in two ways, heads or tails. If a die is thrown it may fall in any one of six ways. Hence, according to the second law, the number of results which can be obtained by tossing first the penny and then the die is $2 \cdot 6 = 12$. This result can be checked by listing the separate possibilities. They are

1. heads and ace
2. heads and deuce
3. heads and three
4. heads and four
5. heads and five
6. heads and six
7. tails and ace
8. tails and deuce
9. tails and three
10. tails and four
11. tails and five
12. tails and six

As a second illustration, consider the modified checkerboard shown in Fig. 1, and ask: In how many ways can a man be moved from the top row, and thereafter a man from the middle row? It is immediately obvious that every man in the top row has two possible moves, which makes a total of 8 ways of moving a man in the top row. After moving the man in the top row there are always two moves possible from the middle row. Hence, the number of possible combinations of moves is $8 \cdot 2 = 16$. Again the combinations of moves can be listed, and will be found to check this result.

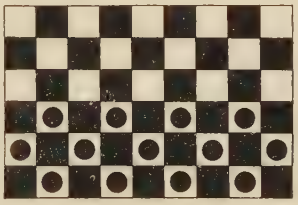


Fig. 1.

Both these illustrations agree with the second law. There are, however, two essential differences between them. In the first illustration the result was obtained by multiplying the total number of possible ways in which a penny may fall by the total number of ways in which a die may fall. In the second illustration, on the other hand, the total number of different moves which are possible from the middle row is 8, one for each of the end men and two for the others; but the correct answer is not obtained by multiplying this number by the 8 possible moves in the top row.

The second difference lies in the fact that in tossing a die and a penny it makes no difference which one is tossed first.

The number of possible combinations is the same in either case. In the checkerboard problem the number of ways in which a man can be moved from the middle row and thereafter a man from the top row is zero, since a man cannot be moved from the middle row at all until a space is opened up for him.

The cause of these differences lies in the fact that the events which take place in the first illustration are *independent*, while those which take place in the second illustration are not. The way a penny falls exerts no conceivable influence over what the die may do; but the way in which the man is moved from the first row determines which men are released in the second row and what moves they may make. The necessity of taking account of such dependence reveals itself in the presence of the words "thereafter" and "in this order" in the statement of the second law.

PROBLEMS

1. Regarding the alphabet as consisting of 21 consonants and 5 vowels, how many distinct five-letter words are possible, each having three consonants and two vowels alternated?

2. In how many of the words of Problem 1 does no letter occur more than once?

3. For purposes of cable code, where a different charge is made according as the combinations of letters are pronounceable or unpronounceable, it might be desirable to obtain a very large number of words of the sort mentioned in Problem 1. How many vowels should an alphabet of 26 letters have, to be most suitable for this purpose?

4. The Greek alphabet has only 24 letters: 17 consonants and 7 vowels. Is it better or worse than the English for the purpose of Problem 3?

§ 8. *Definitions*

From the standpoint of the study of permutations and combinations, a group of objects has three characteristics: the kinds of objects included in the group, the number of

objects of each kind, and the way in which they are arranged. Thus in the group of letters

a b a a,

the fact that there are two kinds of objects, that there are three of one kind and one of the other, and the way in which these are arranged in the group would be the pertinent information.

Two groups of objects are said to form different "combinations" if they differ in the number of any kind of object included.

Thus the combination

a b a a

differs from the combination

a b a,

because the number of **a**'s is not the same as before. It also differs from the combination

a b a b

for the same reason, although the total number of objects in the group in this case is left unchanged.

On the other hand, the combinations

a b a a,
a a b a,
b a a a,
a a a b,

are all the same, since the number of **a**'s and the number of **b**'s is the same in each case.

Two groups of objects are said to form different "permutations" in either of two cases: (a) if they form different combinations; (b) if they form identical combinations but differ in arrangement.

Thus

a b c d,
a b d c,
d c a b,

are identical combinations because each of them has one **a**, one **b**, one **c** and one **d**. They are all different permutations, however, because the arrangement of the letters is different in each group.

The groups

a b c d,
a b d,

are different combinations because they have different numbers of **c**'s. They are therefore also different permutations. As another illustration

a b c d,
a b c e,

are different combinations and therefore also different permutations.

§ 9. *Application of the General Laws of Composition to Permutations and Combinations; Some Typical Examples*

The two general laws stated in § 7 make possible the solution of many of the problems of permutations and combinations. A few examples will illustrate the method of procedure.

EXAMPLE 1.—*How many permutations of three letters each can be formed from the letters* a b c d*?*

The answer to this question may be obtained by thinking of the process involved in writing down the various permutations. In writing down any permutation there are just four ways in which the first letter can occur. After this event has taken place there are three ways of writing the second letter. Finally, the third letter can be written in only two ways. Hence the number of ways of writing three letters is $4 \cdot 3 \cdot 2 = 24$. There is no other conceivable way of putting the letters together. Hence there are just 24 permutations of four letters three at a time.

Table I shows these permutations arranged in the order in which they were supposed to be obtained. The four ways of

writing the first letter correspond to the four vertical columns. In each of these columns there are three pairs corresponding to the three ways of choosing the second letter. Finally, the two members of each pair correspond to the two ways in which the last letter may appear.

EXAMPLE 2.—*How many combinations of three letters each can be formed of the letters* a b c d *?*

TABLE I

PERMUTATIONS OF FOUR LETTERS THREE AT A TIME

a b c	b a c	c a b	d a b
a b d	b a d	c a d	d a c
a c b	b c a	c b a	d b a
a c d	b c d	c b d	d b c
a d b	b d a	c d a	d c a
a d c	b d c	c d b	d c b

Because of the simple numbers involved the easiest way of getting the answer to this problem does not involve the two general laws at all. Whenever a group of three is chosen from four letters, one is left; and it is evident that there will be as many different combinations of *three* letters as there are different ways of having *one* letter left over; that is, four.

Unfortunately, most problems cannot be so easily solved. To illustrate the general process the same answer will be obtained in a less direct way.

Two permutations differ, (*a*) when they are different combinations and (*b*) when they are different arrangements (i.e., permutations) of the same combination. This is a matter of definition. Suppose x is the number of combinations of four things three at a time. Each of these combinations is a group of three, and is capable of a certain number of different arrangements. Call this number y. Every other combination is also capable of y permutations, thus making xy in all. Now it is obvious by definition that no two of these

are identical permutations; and it may easily be seen that every possible permutation is included, since any possible permutation is a particular arrangement of some combination, all of which have been counted. But it is already known that there are just 24 permutations. Hence

$$xy = 24. \tag{1}$$

As for y itself: The first letter of a group of three can be chosen in three ways, the second in two, and the third in one. Hence $y = 3 \cdot 2 \cdot 1 = 6$. Substituting this value in (1) it is found that $x = 4$. This is the same result as before.

§ 10. *Application of the General Laws of Composition to Permutations; General Theorems*

The processes used in Examples 1 and 2 are perfectly general and make it possible to obtain formulæ for the permutations and combinations of groups of objects.

First consider a group of m different objects, and attempt to find the number of permutations of n objects each which can be made from this group of m.

Since the objects are all distinct, the first one in order can be chosen in m ways. Thereafter the second can be chosen in $m - 1$ ways, the third in $m - 2$ ways and so on. In general the number of ways of choosing an object is m minus the number of objects already chosen. When the choice of the last, or nth, object is reached, $n - 1$ will already have been chosen; and therefore the last choice can take place in $m - n + 1$ ways. The entire group of possible choices therefore numbers

$$m(m - 1)(m - 2) \ldots (m - n + 1). \tag{2}$$

This is a perfectly general formula. If $m = 4$ and $n = 3$ the problem treated in Example 1 is obtained. In this case $m - n + 1$ is 2 and the answer is the product of all the integers from 2 to 4 inclusive. This, of course, is the same result as was obtained before.

An important special case of (2) is that in which m and n are equal. In this case $m - n + 1 = 1$ and (2) becomes

$$m(m-1) \ldots 3 \cdot 2 \cdot 1. \tag{3}$$

This is the number of different ways in which a group of m unlike objects can be arranged, or, in technical terms, " the number of permutations of m things m at a time."

A slightly more general problem is: A group contains s kinds of objects. There are m_1 of the first kind, m_2 of the second kind, and so on, the total number of objects being $m = m_1 + m_2 + \ldots + m_s$. In how many distinct ways can they be arranged?

What is meant by saying that m_1 objects are " of the same kind," is that interchanging two of these objects makes no difference in the arrangement of the group. For instance, a a b b is a certain group: it may be thought of as a row of alphabet blocks. If the blocks containing the a's are interchanged the group is still a a b b, and is unaltered.

Suppose the answer to the problem is x. Let one of these x permutations be chosen, and the objects of like kind be tagged to establish their identity. By this means all the objects are rendered distinct. Then the m_1 objects of the first kind will be capable of $m_1(m_1 - 1) \ldots 2 \cdot 1$ permutations among themselves, none of which, however, would have been different from the one chosen if the objects had not been tagged. Those of the second kind will likewise be capable of $m_2(m_2 - 1) \ldots 2 \cdot 1$ permutations among themselves, any of which may be associated with any of the permutations of the objects of the first kind without altering the original permutation of the untagged objects. A similar statement can obviously be made for each of the s kinds of objects. Hence for each of the original permutations there is now a total of

$$(1 \cdot 2 \ldots m_1)(1 \cdot 2 \ldots m_2)(1 \cdot 2 \ldots m_3) \ldots (1 \cdot 2 \ldots m_s)$$

permutations, all of which were made possible by tagging the objects. This makes a total of

$$(1 \cdot 2 \ldots m_1)(1 \cdot 2 \ldots m_2)(1 \cdot 2 \ldots m_3) \ldots (1 \cdot 2 \ldots m_s)x. \tag{4}$$

It must now be noted that every possible permutation of the $m = m_1 + m_2 + \ldots + m_s$ tagged objects is included in this way: for if there were another one, and if the objects were arranged in this permutation and the tags taken off, the result would of necessity be one of the untagged permutations. All possible arrangements which can be made from these, however, have already been provided for in (4). Hence this "other one" must be identical with one of those already provided for.

Finally, it is already known from (3) that the number of permutations of the tagged objects is

$$1 \cdot 2 \cdot 3 \ldots m.$$

Hence, from (4)

$$x = \frac{1 \cdot 2 \cdot 3 \ldots m}{(1 \cdot 2 \ldots m_1)(1 \cdot 2 \ldots m_2) \ldots (1 \cdot 2 \ldots m_s)}. \qquad (5)$$

PROBLEMS

1. How many permutations of the letters of the word "concatenation" are possible?

2. A firm has four positions available, and a list of eleven applicants. How many possible ways are there of filling them?

3. A horseshoe contains eight nails. In how many different orders may they be driven? If enough horses are to be provided so that one shoe may be attached in every possible way, how many miles of four-foot stalls would be required to accommodate them?

4. There are available m_1 objects of one kind, m_2 of another, and so on. How many possible permutations can be built up using n_1 of the first kind, n_2 of the second, and so on? It is assumed that each m is at least as big as the corresponding n.

5. What is the answer to Problem 4 if one n, say, n_1, exceeds the corresponding m?

§ 11. *Factorials; the Gamma Function*

The combination of numbers which presents itself in (3) is a very common one in more than one branch of mathematics; so much so that it has been given a name and a shorthand symbol. It is called

"factorial m," and is usually written as $m!$ in modern books. Some years ago a more common notation was $\lfloor m$.

Factorial m may therefore be defined as the product of all integers from 1 to m, inclusive. When so defined, it has no meaning unless m itself is an integer; but this defect can easily be remedied by certain artifices.

When m is an integer, it is easily shown by direct integration that

$$\int_0^\infty x^m e^{-x} dx = m!. \tag{6}$$

As yet, when m is fractional, $m!$ means nothing, and it is therefore impossible to say that the equation (6) is either true or false. If, however, (6) instead of (3) is adopted as the *definition* of $m!$, the symbol will have the same significance as before for integers, and it will also have a meaning for fractions. This is the definition usually adopted.[1]

Even this definition applies only when $m > -1$, for the integral does not have a value for other values of m. The integral obeys a certain law, however, by means of which the definition can be extended to all real numbers.

Integrating (6) by parts, it is found that

$$\int_0^\infty x^m e^{-x} dx = -x^m e^{-x}\Big|_0^\infty + m\int_0^\infty x^{m-1} e^{-x} dx.$$

As the second term vanishes at both limits of integration[2] this equation may be rewritten

$$m! = m(m-1)!. \tag{7}$$

[1] As a matter of fact, a much more complicated definition is given for purely mathematical purposes, in order to overcome the limitation to which the next paragraph of text calls attention. So far as these notes are concerned, however, the definition (6) is entirely satisfactory, and the student will not be misled by it.

[2] The second term obviously vanishes when $x = 0$. When $x = \infty$ it takes the form $\infty \cdot 0$, which is indeterminate. To evaluate this "indeterminate form," the usual process is to replace $x^m e^{-x}$ by x^m/e^x which is identical with it, but takes the form ∞/∞. The next step is to differentiate numerator and denominator separately, getting

$$\frac{mx^{m-1}}{e^x}.$$

This is still in the form ∞/∞, so another differentiation is performed, and then another and so on. The denominator in each successive step remains the same, but each step

This law can be more easily obtained from (3) when m is integral; having been obtained from (6) it is true whether m is integral or not. Moreover, if it is taken as a universal property of $m!$ it makes it possible to define $m!$ when m is negative.

For instance: it is known [1] that $\frac{1}{2}! = \frac{1}{2}\sqrt{\pi}$. By setting $m = \frac{1}{2}$

---

reduces the degree of the numerator by 1. After n differentiations the form appears

$$\frac{m(m-1)(m-2)\ldots(m-n+1)x^{m-n}}{e^x}.$$

If m is an integer, by setting $n = m$ the numerator becomes a finite constant and the fraction (for $x = \infty$) takes the form $C/\infty = 0$.

If m is a fraction, n can be chosen as the next *larger* integer, in which case the fraction takes the form

$$\frac{C}{x^{n-m}e^x}.$$

As $n - m$ is positive, *both* terms in the denominator become infinite with x, and *a fortiori* their product does. Hence, as before, the fraction takes the form $C/\infty = 0$.

Those who wish to renew their acquaintance with this process may look up the subject of "indeterminate forms" in any Calculus textbook; e.g., March and Wolf, *Calculus*, pp. 324–327.

[1] By definition

$$\tfrac{1}{2}! = \int_0^\infty e^{-x}x^{\frac{1}{2}}\,dx.$$

Replace x by y^2. Then

$$\tfrac{1}{2}! = 2\int_0^\infty e^{-y^2}y^2\,dy.$$

But it makes no difference whether the variable is y or z. Hence

$$\tfrac{1}{2}! = 2\int_0^\infty e^{-z^2}z^2\,dz.$$

Now, multiply these equations member by member and place the z's under the sign of y-integration, with respect to which they are constant.

$$(\tfrac{1}{2}!)^2 = 4\int_0^\infty\int_0^\infty e^{-(y^2+z^2)}y^2z^2\,dy\,dz.$$

Finally, note that in the yz-plane $dy\,dz$ is the element of area, and that the integral extends over the entire first quadrant. Rewriting the integral in terms of polar coördinates, which means (see § 68)

$$y = r\cos\theta$$
$$z = r\sin\theta$$
$$dy\,dz = r\,dr\,d\theta,$$

in (7), and inserting this known value, it is found that

$$\tfrac{1}{2}\sqrt{\pi} = \tfrac{1}{2}(-\tfrac{1}{2})!$$

or

$$(-\tfrac{1}{2})! = \sqrt{\pi}.$$

Similarly by putting $m = -\tfrac{1}{2}$,

$$\sqrt{\pi} = -\tfrac{1}{2}(-\tfrac{3}{2})!$$

or

$$(-\tfrac{3}{2})! = -2\sqrt{\pi}.$$

By starting from a suitable positive factorial and using this device it is possible to obtain the factorial of any negative number whatever.

The factorials of negative integers are particularly interesting. It is obvious that $1! = 1$. Setting $m = 1$ in (7) it is found that $0!$ also equals unity. Then putting $m = 0$ in (7)

$$0! = 0(-1)!$$

or

$$(-1)! = \tfrac{1}{0} = \infty.$$

By the same process:

$$(-2)! = \frac{(-1)!}{-1} = \infty,$$

$$(-3)! = \frac{(-2)!}{-2} = \frac{(-1)!}{(-1)(-2)} = \infty,$$

$$(-4)! = \frac{(-3)!}{-3} = \frac{(-1)!}{(-1)(-2)(-3)} = \infty,$$

---

it becomes

$$(\tfrac{1}{2}!)^2 = 4\int_0^\infty dr \int_0^{\frac{\pi}{2}} d\theta\, r^5 e^{-r^2} \cos^2\theta \sin^2\theta.$$

But

$$\int_0^{\frac{\pi}{2}} \cos^2\theta \sin^2\theta\, d\theta = \pi/16,$$

and

$$\int_0^\infty r^5 e^{-r^2} dr = 1.$$

Hence

$$\tfrac{1}{2}! = \frac{\sqrt{\pi}}{2}.$$

and so on. Since infinity divided by any succession of finite quantities is still infinite, it follows that *the factorial of any negative integer is infinite.*

The behavior of the function will be made clearer by reference to Fig. 2, in which the factorials of numbers between -6 and $+3$ are plotted.

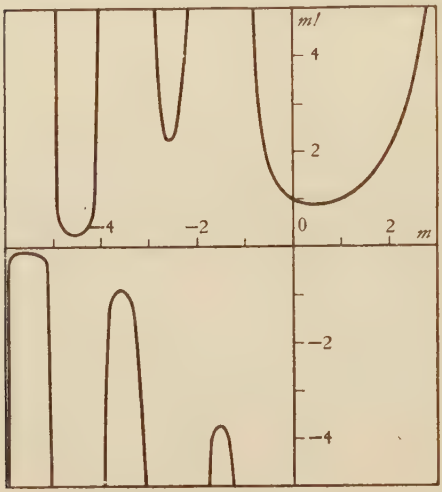


Fig. 2.—The Factorial.

For completeness one more fact should be mentioned. It is customary among mathematicians to speak of "factorials" only when referring to the factorials of positive integers. In other cases they talk about a "gamma-function" or "Γ-function." This function is related to the factorial discussed above by the law

$$\Gamma(m) = (m - 1)!.$$

Thus $\Gamma(3) = 2! = 2.1 = 2$; and $\Gamma(0) = (-1)! = \infty$. As the multiplication of names and symbols serves no useful purpose, all such numbers will be called "factorials" in what follows.

From the above discussion the student should retain the following facts:

(*a*) Factorial m, when m is an integer, means the product of the integers from 1 to m. It is written $m!$ or $\lfloor m$.

(*b*) The symbol has a meaning for other numbers than integers. Its value may be found tabulated, just as logarithms are.

(*c*) The factorial of zero is equal to unity. [$0! = 1$.]

(*d*) The factorial of a negative integer is infinite. [$(-j)! = \infty$ if $j > 0$.]

(*e*) The factorials of negative *fractional* numbers are not infinite.

There will be many occasions to use factorials in what follows, but they will generally be the factorials of integers. To simplify numerical calculations, the values of the factorials of integers up to 200 are listed in Appendix I. As many of the numbers are inconceivably large, only a few significant figures are written, followed by

the power of ten by which they must be multiplied. Thus 20! is given as $2.432\ 9020^{18}$ which means

$$2.432\ 9020 \cdot 10^{18} = 2{,}432{,}902{,}000{,}000{,}000{,}000.$$

In Appendix II are given the logarithms of the factorials up to 1200.

§ 12. *Restatement of General Theorems in Permutations*

Let the symbol P_n^m stand for the number of permutations of m things n at a time. Then (2) may be rewritten as

$$P_n^m = m(m-1)\ldots(m-n+1)$$

$$= \frac{m(m-1)\ldots(m-n+1)(m-n)(m-n-1)\ldots 2\cdot 1}{(m-n)(m-n-1)\ldots 2\cdot 1}.$$

Making use of the factorial notation this reduces to

$$P_n^m = \frac{m!}{(m-n)!}. \tag{2}$$

Similarly (3), which represents the number of permutations of m objects m at a time, becomes

$$P_m^m = m!. \tag{3}$$

Finally (5), which gives the number of possible arrangements of a group composed of s kinds of objects, m_1 of the first kind, m_2 of the second, and so on, is

$$P_{m_1, m_2, \ldots, m_s}^{m_1, m_2, \ldots, m_s} = \frac{m!}{m_1!\, m_2! \ldots m_s!}. \tag{5}$$

§ 13. *Application of the General Laws to Combinations*

The number of combinations of m things taken n at a time can be computed by the same line of argument as was used in Example 2 of § 9. Each combination is capable of $n!$ permutations within itself. If, therefore, the number of combinations is denoted by C_n^m, the total number of permutations will be

$$P_n^m = C_n^m\, n!.$$

The value of P_n^m, however, is known from (2). Hence

$$C_n^m = \frac{m!}{n!\,(m-n)!}. \tag{8}$$

This is the general formula for the number of combinations of m things taken n at a time. In particular, if m is 4 and n is 3, the case is the same as that considered in Example 2 of § 9. Substituting these values in (8) the answer is found to be

$$\frac{4!}{3!\,1!} = \frac{4\cdot 3\cdot 2\cdot 1}{3\cdot 2\cdot 1\cdot 1} = 4,$$

which checks the result obtained in § 9.

Some further examples may be given.

EXAMPLE 3.—*How many straight lines can be drawn through five points in such a way that each line contains two points?*

A straight line can be drawn through any pair of points whatever. There will therefore be as many straight lines as there are pairs of points. The only question which remains for consideration is as to whether the pairs of points p_1p_2 and p_2p_1 are to be considered as identical or distinct. In the former case the problem becomes a problem in *combinations*; in the latter it is a problem in *permutations*.

It is obviously possible to draw a line from a to b and also one from b to a, but in the sense in which lines are usually thought of these two lines would be identical. In other words, the direction in which the pencil travels in marking out the line is of no consequence. Hence p_1p_2 and p_2p_1 must be thought of as identical groups and the problem becomes a problem in combinations. Its solution is therefore [1]

$$C_2^5 = \frac{5!}{2!\,3!} = 10.$$

[1] There are exceptions to this answer. For instance, if all five points lie upon the X-axis of a system of co-ordinates there are not ten lines which can be drawn through them. Instead the X-axis itself is the only possible line. In obtaining the answer above it is tacitly assumed that no three points are collinear.

EXAMPLE 4.—*How many distinct hands of 13 cards can be dealt from a full pack without a joker?*

This is obviously a question of combinations since the order in which the cards are dealt is immaterial. The answer is

$$C_{13}^{52} = \frac{52!}{13!\,39!} = 635{,}013{,}559{,}600.$$

PROBLEMS

1. How many possible stacks of 7 dominoes can be dealt from a set of 28?

2. There are 11 candidates from among whom 200 electors are to choose a board of directors. Half of the electors have two votes each to one each for the rest. The board is to consist of five members. On each ballot the lowest man is discarded, until only five remain. How many boards are possible?

§ 14. *Some General Properties of C_n^m; Pascal's Triangle*

The first solution of Example 2, § 9, was obtained by noting that whenever a different combination of three letters was taken out of the four supplied, a different combination of one was left. A little reflection shows that this is a general law; whenever a different combination of n things is taken from a group of m a different combination of $m - n$ things remains. There must therefore be at least as many distinct combinations of $m - n$ as there are of n, that is,

$$C_{m-n}^m \geqq C_n^m. \tag{9}$$

This rule, however, works backward just as well as forward. Whenever a different group of $m - n$ is taken out, a different group of n remains. Hence, the distinct groups of n must be at least as numerous as the distinct groups of $m - n$. This leads to

$$C_n^m \geqq C_{m-n}^m. \tag{10}$$

By comparing (9) and (10) it is seen that *both* can be true

only when the equality signs are used. Hence it follows as a general rule that [1]

$$C_n^m = C_{m-n}^m. \tag{11}$$

Another general property of the quantities C_n^m can be obtained from the obvious equation

$$(m - n)(m - 1)! + n(m - 1)! = m!.$$

If this equation is divided throughout by both $n!$ and $(m - n)!$ and a few obvious cancellations are made, it reduces to

$$C_n^{m-1} + C_{n-1}^{m-1} = C_n^m. \tag{12}$$

This equation belongs to the class known as "recursion formulæ." In other words, it is of such a nature that once the combinations that can be made from a group of $m - 1$ things are all known it is possible to find the combinations that can be made from m things. For instance, in the column headed 5 in Table II, 5 is the number of combinations of one thing each which can be made from a group of five distinct objects; 10 is the number of combinations of two each; and so on, as indicated by the marginal numbers.[2]

Similarly the column headed 6 contains the values of C_n^6. It can be obtained from the first column by the use of (12). For upon putting m equal to 6, (12) becomes

$$C_n^6 = C_n^5 + C_{n-1}^5,$$

[1] This rule can also be derived as an immediate consequence of (8), for

$$C_{m-n}^m = \frac{m!}{(m-n)![m-(m-n)]!} = \frac{m!}{(m-n)!\,n!} = C_n^m.$$

The logical argument, however, may carry somewhat greater conviction than the formal manipulation of symbols because it makes it possible to "see why" the answer is correct.

[2] There is also added, opposite the marginal number 0, the value taken by (8) when $n = 0$. This value is unity, regardless of the value of m, so that the entire row is filled with 1's. Of course, this quantity does not make sense when phrased as "the number of different groups of no things each which can be formed from a group of m things"; it must be regarded as an extension of the meaning of symbol C_n^m by definition. It is required for the general validity of (11) and (12).

which says that each entry in the column headed 6 is the sum of the adjoining number in the 5 column, and the number next above the latter. Thus C_1^6 is the sum of 5 and 1; C_2^6 is the sum of 10 and 5, and so on.

The 7 column can be obtained in the same way from the 6 column.

TABLE II

VALUES OF C_n^m

n	$m = 5$	$m = 6$	$m = 7$
0	1	1	1
1	5	6	7
2	10	15	21
3	10	20	35
4	5	15	35
5	1	6	21
6	0	1	7
7	0	0	1

Appendix III contains the values of C_n^m for groups not larger than 100. Because of the large numbers involved this table is written in the same notation as that which was used in Appendix I. Moreover, the size of the table has been reduced by making use of the property (11), according to which the numbers in any column repeat themselves in the inverse order after the center of the column is passed. For this reason it is only necessary to tabulate the numbers up to the point where they begin repeating. For instance, C_{31}^{50} is not given in the table, but is known from (11) to be equal to C_{19}^{50}, which is given as 3.0405943^{13}.

§ 15. *Some General Properties of* C_n^m*; The Binomial Theorem*

The rule for algebraic multiplication is, that every term of the one factor of the product is to be multiplied by every term of the other factor of the product, and all of these partial products are then to be added together. This is what mathe-

maticians call the "distributive law of multiplication." It can also be phrased in the following words: "The product of two polynomials is the sum of the partial products of all possible combinations of terms obtained by taking one term from each of the factors."

Suppose a product of two factors has been obtained in this way and that this product is then multiplied by a third polynominal. The product thus obtained contains the sum of all possible partial products of three terms each, which can be formed by taking one term from each of the three polynomials.

In general, when m factors are involved the product must contain the sum of every possible partial product which can be formed by taking one term from each of the m polynomials.

Suppose now that each of the m factors is taken as the same binomial $x + y$. The product will then represent $(x + y)^m$. Since there are only two possible terms, x and y, every partial product must be a power of x multiplied by a power of y. Furthermore, since one term must be taken from each factor, the sum of the two powers must always be equal to m. In other words, the product, when it has been completely worked out, must take the form

$$c_0 x^m + c_1 x^{m-1} y + c_2 x^{m-2} y^2 + \ldots + c_m y^m, \qquad (13)$$

where the values of the c's are as yet unknown.

However, it is not difficult to determine the numerical values of these c's, for the complete product must contain just as many partial products of the form $x^n y^{m-n}$ as there are different ways of choosing x's from n different factors, the order of choice being immaterial; in other words, it is C_n^m. The y's, of course, come from the remaining $m - n$ factors. That is, the coefficient of the nth power of x in the binomial expansion of $(x + y)^m$ is equal to the number of combinations of m things taken n at a time.

Substituting this value in (13) it takes the form

$$C_0^m x^m + C_1^m x^{m-1} y + C_2^m x^{m-2} y^2 + \ldots + C_m^m y^m, \qquad (14)$$

or in shorthand notation [1]

$$(x+y)^m = \sum_{n=0}^{m} C_n^m x^{m-n} y^n. \qquad (15)$$

Referring back to Appendix III the reader will at once recognize the sequences of numbers

1, 1
1, 2, 1
1, 3, 3, 1

as belonging to the well-known binomial expansions

$$(x+y)^1 = 1\cdot x + 1\cdot y$$
$$(x+y)^2 = 1\cdot x^2 + 2\cdot xy + 1\cdot y^2$$
$$(x+y)^3 = 1\cdot x^3 + 3\cdot x^2 y + 3\cdot xy^2 + 1\cdot y^3.$$

Even the integer 1 standing alone under the heading $m = 0$ fits into this arrangement because $(x+y)^0 = 1$.

---

[1] The Σ-notation for the sum of a number of terms all of which follow the same law of formation is very convenient and enables us to write quite complicated expressions in a comparatively simple form. This notation will be frequently used in what follows. The idea upon which it is based can be easily illustrated by reference to (15).

The product

$$C_n^m x^{m-n} y^n$$

is called the "general term" of (14). This means that by assigning different values to n it is possible to obtain every term of (14). For instance if n is put equal to 2 the term

$$C_2^m x^{m-2} y^2$$

is obtained. If n is put equal to 1 the term

$$C_1^m x^{m-1} y$$

is obtained. Proceeding in this fashion and taking the proper set of values for n, every term of (14) may be duplicated.

The Σ of (15) is a command to add together the proper set of terms of this sort. The symbols written above and below it define "the proper set" by telling the largest and smallest values of n respectively. It is understood that all intermediate integers are to be used.

Thus the right-hand side of (15) is shorthand notation for the instruction "build up a term of the form

$$C_n^m x^{m-n} y^n$$

for every integral value of n from 0 to m inclusive, and add all these terms together."

§ 16. *The Solution of More Complicated Problems*

There are many problems in permutations and combinations which do not fall into the simple classes dealt with above. Such problems can often be solved by the intelligent use of the general laws stated in § 6. A few examples will illustrate the general line of attack.

EXAMPLE 5.—*How many combinations of four letters each can be made from the word* pepper*?*

This example is complicated by the fact that some letters are repeated. The simplest method of dealing with it is to note that there are two general classes of combinations, those which contain **r** and those which do not.

Those which contain **r** must also contain some combination of three formed from three **p**'s and two **e**'s. It is a simple matter to list these combinations, which turn out to be

p p p,
p p e,
p e e.

Those groups which do not contain **r** must have four letters chosen from the three **p**'s and the two **e**'s. It is evident that there are only two such combinations:

p p p e,
p p e e.

Since the cases which contain **r** and those which do not are mutually exclusive, the first general law of composition gives the total number of combinations as five.

In solving this example it was not necessary to make use of any of the formulæ which have been derived for permutations and combinations. An illustration in which those formulæ are useful as an adjunct to the general laws is the following:

EXAMPLE 6.—*How many combinations of four letters each can be made from the letters of the word* provocative*?*

This word has eleven letters, two **o**'s, two **v**'s, and one each of the seven letters **a, c, e, i, p, r, t.**

The possible combinations of the **o**'s and **v**'s are

oovv	oov	oo	o
	ovv	ov	v
		vv	

The first column contains a combination of four; it is therefore one of the cases sought. The second column contains combinations of three, and by the second law of composition it is known that each of these can be combined with every possible combination of one formed from the seven distinct characters. Likewise each of the combinations in the third column can be combined with the C_2^7 combinations of 2, and each of the fourth column with the C_3^7 combinations of 3. Finally there are C_4^7 combinations of four letters containing neither **o**'s nor **v**'s. Using the first law, the total number of combinations is found to be

$$1 + 2 \cdot C_1^7 + 3 \cdot C_2^7 + 2 \cdot C_3^7 + C_4^7 .$$

This works out to be 183 combinations in all.

It is frequently necessary in considering problems in permutations and combinations to make use of artifices of this sort. In such cases the ease with which the solution is obtained depends very largely upon the appropriateness of the method of attack, so that experience and that sense which we call intuition are necessary before such problems can be satisfactorily handled.

§ 17. *A Complicated Problem in Permutations*

The following problem is introduced for two reasons: In the first place it leads to a result which is of some interest on its own account; in the second it affords an excellent drill in the purely formal thinking that is often associated with the use of the Σ notation.

It deals with a question similar to that treated in Examples 5 and 6, except that *permutations* instead of *combinations* are asked for. It may be stated as follows:

EXAMPLE 7.—*How many different permutations of n things each can be made from m_1 of one kind, m_2 of a second, and so on, there being s different kinds available? No restriction is placed upon the proportion of each kind to be included in the permutations.*

Now, though the exact make-up of any permutation is unknown, it must have *some* number of objects of each kind. Let them be denoted by $n_1, n_2, \ldots, n_s$. Then according to the result of Problem 4, § 10, there are just

$$P_{n_1, n_2, \ldots, n_s}^{m_1, m_2, \ldots, m_s} = \frac{n!}{n_1!\, n_2! \ldots n_s!} \tag{16}$$

permutations for which the n's remain unchanged. Obviously the total number of permutations, $P_n^{m_1, m_2, \ldots, m_s}$, will be obtained by adding together terms of this sort, one term for every possible way in which a group of n may be made up. The result must therefore take the form

$$P_n^{m_1, m_2, \ldots, m_s} = \sum_{n_1} \sum_{n_2} \cdots \sum_{n_s} P_{n_1, n_2, \ldots, n_s}^{m_1, m_2, \ldots, m_s},$$

provided the proper limits are assigned to the summations.

These limits are found as follows:

In the first place, since

$$n_1 + n_2 + \ldots + n_s = n,$$

one of the n's, say n_s, is not arbitrary, but has its value fixed as soon as the others are known. The n_s-summation therefore consists of but a single term for which

$$n_s = n - n_1 - n_2 - \ldots - n_{s-1}.$$

But if there is but a single term the sign of summation may be dropped. This gives

$$P_n^{m_1, m_2, \ldots, m_s} = \sum_{n_1} \sum_{n_2} \cdots \sum_{n_{s-1}} \frac{n!}{n_1!\, n_2! \ldots n_{s-1}!\, (n - n_1 - n_2 - \ldots - n_{s-1})!}. \tag{17}$$

Next, it may be noted that n_1 cannot exceed m_1, n_2 cannot exceed m_2, and so on. This suggests that the upper limits of summation shall be $m_1, m_2, \ldots, m_{s-1}$. But caution must be observed in this matter. For if some one m is greater than n this would seem to require the use of permutations in which the number of objects of one kind exceeded the total number of all kinds, which is obviously impossible.

However, if any n_j, say n_1, exceeds n, the term

$$(n - n_1 - n_2 - \ldots - n_{s-1})!$$

becomes the factorial of a negative integer, which is known to be infinite. In fact this occurs whenever the sum of $n_1 + n_2 + \ldots + n_{s-1}$ exceeds n. As this factorial occurs in the denominator of (17), it follows that whenever a set of values of $n_1, n_2, \ldots, n_s$ is used, which is impossible because the total exceeds n, the corresponding term in (17) automatically becomes zero. Hence, even though $m_1, m_2, \ldots, m_{s-1}$ may be bigger than the *correct* limits, they may nevertheless be used, for the additional terms which they introduce into (17) are all zeros.

Finally the lower limits of summation must be obtained. To this end, consider first the case of n_1, and think of a particular illustration in which $n = 7$ and the groups of like objects are nine **a**'s, two **b**'s and three **c**'s: that is, $m_1 = 9$, $m_2 = 2$ and $m_3 = 3$. Obviously, since there cannot be more than two **b**'s and three **c**'s in any permutation there must be at least two **a**'s. That is, n_1 cannot be less than 2.

By the same line of argument, since $n_2, n_3, \ldots, n_s$ cannot exceed $m_2, m_3, \ldots, m_s$, respectively, it follows that their sum cannot exceed $m_2 + m_3 + \ldots + m_s$. But as the total number of objects in a permutation is fixed at n, the smallest number of objects of the first kind will occur in those permutations which have the largest aggregate number of objects of all other kinds. That is, n_1 cannot be smaller than

$$n - m_2 - m_3 - \ldots - m_s. \qquad (18)$$

The same argument applies to n_2, except that, in carrying out the second summation the value of n_1 is supposed to be known, and therefore only the remaining n's must be made as large as possible. Thus n_2 cannot be smaller than

$$n - n_1 - m_3 - \ldots - m_s.$$

Similarly the other n's must satisfy the following inequalities:

$$n_3 \geqq n - n_1 - n_2 - m_4 - \ldots - m_s,$$

$$n_4 \geqq n - n_1 - n_2 - n_3 - m_5 - \ldots - m_s,$$

$$\ldots \qquad \ldots \qquad \ldots$$

$$n_{s-1} \geqq n - n_1 - n_2 - \ldots - n_{s-2} - m_s.$$

Before these numbers can be used as limits of summation, however, it must be observed that $n_1, n_2, n_3, \ldots, n_s$ cannot be less than zero.

Hence the lower limit of summation on n_1 is (18) when that is positive: otherwise it is zero; and similar statements must be made for the other n's. However, if (18) were negative, and if it were used as a lower limit of summation, it would introduce certain excess terms for which n_1 was negative. Due to the presence of $n_1!$ in the denominator of (17) all these terms would be zero. Hence the use of (18) would give a correct result.

A like statement applies for every other summation. Hence

$$P_n^{m_1, m_2, \ldots, m_s} = \sum_{\substack{n_1 = n - m_2 \\ - \ldots - m_s}}^{m_1} \sum_{\substack{n_2 = n - n_1 - m_3 \\ - \ldots - m_s}}^{m_2} \cdots \sum_{\substack{n_{s-1} = n - n_1 \\ - \ldots - n_{s-2} - m_s}}^{m_{s-1}} \frac{n!}{n_1!\, n_2! \ldots n_{s-1}!\, (n - n_1 - n_2 - \ldots - n_{s-1})!}. \tag{19}$$

This, then, is the formal solution of the problem.

PROBLEMS

1. In how many ways can four cards in sequence in the same suit be chosen from a full pack if the order of choice is immaterial, so that

six	six	eight
seven	nine	six
eight	eight	seven
nine	seven	nine

are all regarded as identical?

2. In how many ways can a cribbage hand of six cards be dealt, (*a*) if the order of dealing is taken into account, so that identical hands, the cards of which appear in different orders, are regarded as different? (*b*) if the order of dealing is neglected?

3. In how many ways can seven keys be arranged on a ring?

4. A printing telegraph machine contains a number of sliding bars each capable of taking two positions in response to current pulses of two different types. When the bars have all been set, one and only one character is selected for printing. It is evident that the number of characters which the machine is capable of selecting will depend upon the number of bars. How many bars are required to handle 50 characters?

5. If in the printing telegraph of Problem 4 the bars are capable of taking three positions instead of two, how many bars are required?

6. In how many ways can p $+$'s and n $-$'s be placed in a row so that no two $-$'s come together?

7. How many different combinations can be made of 10 objects, 3 of which are alike and 3 others alike, the remaining 4 being different, (*a*) if the number of objects in a combination is not restricted; (*b*) if it is equal to 4?

8. How many different connections must a telephone exchange be capable of setting up, if it accommodates 10,000 subscribers?

The cylinder lock illustrated in Fig. 3 contains five tumblers. These tumblers are in the form of pins, cut in two parts so that when forced into the proper position by the edge of the key (Fig. 3*b*) they offer no restraint to the rotation of the cylinder. The cuts may be made at any one of ten points along the pin. When the key is out, as in Fig. 3*a*, or if the wrong key is inserted, the cuts do not all coincide with the edge of the cylinder, and it cannot move.

If a master-key is required for a number of locks, certain tumblers may be cut in more than one place, as in Fig. 3*c*. When this is done, one set of cuts is the same on the tumblers of all locks; they can

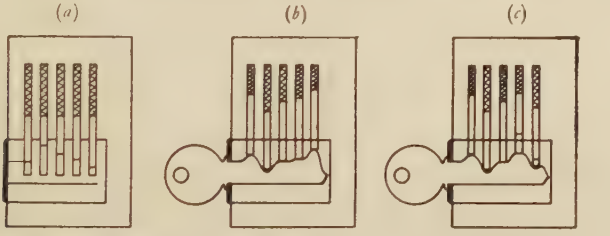


FIG. 3.

therefore all be opened by the same master-key. The other set of cuts is different for every lock, so that the key corresponding to any one will not operate the rest. In Fig. 3*c*, those cuts which are not in line would be brought in line by the key shown in *b*. Thus this particular lock could be operated by *either* key.

The following problems refer to locks constructed on this system:

9. How many different locks can be made without changing the keyway? (No double-cut pins.)

10. When the key of a lock is stolen, and it is desired to protect the owner against possible entry, the pins are taken out and inter-

changed. If three of the pins happen to be cut differently, the remaining two being alike, how many times may this be done without duplication?

11. If double-cut pins are used on the first and fourth tumblers, to provide for a master key, how many distinct locks will it open?

12. Each of five floors of a hotel is to have a separate and distinct master-key. How many rooms may each floor have, without permitting any guest's key to open any but his own room, provided (*a*) only one double-cut pin is used; (*b*) two double-cut pins; (*c*) three double-cut pins?

References for Outside Reading

1. Todhunter: *Algebra*, pp. 286–297.
2. Chrystal: *Algebra*, Vol. II, Chap. XXIII.
3. Whitworth: *Choice and Chance*, Chaps. I and II.
4. Netto: *Lehrbuch der Kombinatorik*, Chaps. I, II, XIII.

CHAPTER III

Elementary Principles of the Theory of Probability

§ 18. *Complementary Probabilities*

The events "A happens" and "A does not happen" are mutually exclusive. Hence, by Convention II, § 3, the probability that "either A happens or A does not happen," is the sum of their separate probabilities. But one or the other of these two events is certain to occur, and therefore the sum must be unity. Hence, if the probability that A happens is denoted by $P(A)$ and the probability that A does not happen by $P(\bar{A})$ it follows that

$$P(A) + P(\bar{A}) = 1,$$

or

$$P(A) = 1 - P(\bar{A}).$$

The numbers $P(A)$ and $P(\bar{A})$, which represent the probability of an event taking place, and the probability of it not taking place, are known as "complementary probabilities."

§19. *Unconditional Probabilities*

The simplest sort of problems in the theory of probability are known as "problems in unconditional probability." Their principal characteristic is the assurance with which the conditions surrounding them can be stated. It is impossible to describe them more exactly, as their classification is of a vague and somewhat illogical nature, but the implication of the name, which is a very useful one, will become clear in the course of a few sections. They can frequently be solved by the

direct application of the fundamental definition of probability. A few examples will illustrate this type of problem.

EXAMPLE 8.—*What is the probability of throwing an ace with a die?*

There are six faces to the die, only one of whch may appear. In an actual die they are not equally likely to appear, for the die is certainly unsymmetrical to a greater or lesser degree. The nature of this asymmetry is unknown, however, and therefore plays no part in the problem. There are other known characteristics of the faces, such as the number and arrangement of the dots, but these characteristics are not pertinent except in so far as they affect the symmetry of the die, and their effect in this respect is unknown. Hence it must be concluded that each of the six faces is "equally likely" to appear. Of this complete group of six faces only one is an ace. Hence, by definition, the chance of an ace appearing is $\frac{1}{6}$.

EXAMPLE 9.—*The letters of the word* tailor *are written on cards. The cards having been thoroughly shuffled, four are drawn in order. What is the probability that the result is* oral *?*

The number of permutations of six distinct things four at a time is $\frac{6!}{2!} = 360$. These permutations differ in no known pertinent respect except their identity. Hence they form a complete group of equally likely events. Only one of these events, however, is the word "oral." Hence the answer to the question is $\frac{1}{360}$.

EXAMPLE 10.—*The letters of the word* pepper *are written on cards. The cards having been thoroughly shuffled, four are drawn in order. What is the chance that the result is* peep *?*

In this case a distinction must be made between *cards* and *letters*. The permutations of the *cards* are all equally likely. The permutations of the *letters* need not be. The permutations of the cards form a complete group of 360 events of which a certain subgroup gives the desired word. The problem is to find the number of permutations in the subgroup.

Any permutation of the three cards bearing the **p**'s leaves the result unchanged so far as letters are concerned. The same is true of the **e**'s. Of course, every permutation of the **p**'s can be combined with every permutation of the **e**'s so that the total number of permutations which leave the group "peep" unchanged is $P_2^3 \, P_2^2 = 6 \times 2 = 12$. The desired subgroup of permutations therefore numbers 12, and the answer is $\frac{12}{360} = \frac{1}{30}$.

EXAMPLE 11.—*First Appearance of the Psychic Research Problem.*[1] — *A spiritualistic medium claims to be able to tell the color of a playing card without seeing it. In order to test her claims an experiment is conducted with four red and four black cards. These cards are thoroughly shuffled and placed face down on the table. The medium is told that there are four red and four black cards, but presumably knows nothing as to their arrangement. The experimenter picks up a card and without either looking at it himself or showing it to the medium asks its color. If she answers "red," he places it at one side of the table. If she answers "black," he places it on the other side of the table. This process is repeated until all cards are exhausted.*

If the medium does not have the ability which she claims to possess, what is the chance that there will be just one black card in the pile that should be red?

The order in which the medium will call her "reds" and "blacks" is, of course, unknown; but if she has no power of detecting the nature of the cards, the order of calling will be quite independent of that in which the cards actually appear. Her chance of success would, in fact, be just the same if she

[1] This example, which recurs in various forms, has the following interesting history: A certain pseudo-scientific hoax, of which the problem is a disguised formulation, was under investigation by a friend of mine. He was anxious to formulate the number of "reds" and "blacks," and other features of the experimental procedure, so as to make the chance of an *accidental* high score, and particularly of ambiguous scores which the "medium" would undoubtedly regard as favorable, as small as possible. It was necessary to have a reasonable proportion of "reds"; and the number of "cards" which the "medium" would consent to handle was limited. We finally arrived at a set-up which, while none too satisfactory, was the best we could get; and with fear and trepidation my friend conducted the experiment. The outcome justified his fears — one of the least probable results occurred. The "medium" got the lowest possible score.

were to state the order in which she expected the cards to appear before they were dealt—that is, if she claimed the gift of "prophecy" instead of "occult understanding." Suppose, then, that we think of this order in which she calls the cards, whatever it is, as a "standard order," and ask for the chance that the cards are so dealt as to match this standard order to just the extent prescribed by the statement of the problem.

It is asserted that the medium knows that there are just four red and four black cards. It may therefore be accepted as a fact that she will call red and black exactly four times each; that is, the standard order contains four reds and four blacks.

Now turning attention to the order in which the cards appear, it is at once obvious that the $P_{4,4}^{4,4} = 70$ possible permutations constitute a complete group of equally likely and mutually exclusive events. Among these there is a certain subgroup which matches the standard order in exactly three reds. If the number of events contained in this subgroup can be found, the problem will have been solved.

This can be done by thinking of the process of laying down a permutation beside the standard order in such a way as to satisfy the conditions of the problem. To start with, a black card can be laid down in any one of the four red positions of the standard order. After this has been done the remaining red positions must all be filled by red cards. Then the red card which remains can be placed in any one of the four positions which are occupied by black cards in the standard order, after which the remaining positions can only be filled with black cards. There are therefore exactly $4 \cdot 4 = 16$ permutations which satisfy the conditions of the problem. The answer to the problem is therefore $\frac{16}{70}$ or $\frac{8}{35}$.

PROBLEMS

1. A connector switch in a step-by-step exchange reaches ten subscribers on each of ten levels — 100 subscribers in all. Every subscriber is represented on such a switch. What is the chance that Mr. A's line appears on the third level?

2. Assume that the various operations of a connector switch consume time as follows:

(*a*) First vertical step	0.15 sec.
(*b*) Succeeding vertical steps	0.10 sec.
(*c*) First horizontal step	0.25 sec.
(*d*) Succeeding horizontal steps	0.10 sec.

An observer visiting the exchange watches a connector until it operates, and notes the operating time. What is the probability that it is less than 0.66 sec.? What is the probability that it lies between 0.66 and 1.50?

3. Mr. A makes two calls during an hour and is twice called by other subscribers, of whom one is Mr. B. The calling subscribers, in case they find Mr. A busy, repeat their calls. Each call occupies one of the 60 minutes of the hour. What is the chance that Mr. B is successful on his first attempt to call?

4. In the psychic research experiment of Example 11, what is the chance of the medium scoring 100 per cent.

5. If, in the psychic research experiment of Example 11, six red cards and two black cards are used, what is the chance of the medium scoring 100 per cent?

6. If, in the psychic research experiment of Example 11, six red cards and two black cards are used, what is the chance that the medium places just one wrong in each color?

7. A milkman starts on his route with ten dozen quarts of fresh milk, together with five dozen quarts left over from the preceding day. Having delivered 100 quarts, he arrives at the home of Mrs. A, who receives one quart. What is the chance that it is stale?

8. A batch of one thousand lamps is five per cent bad. If five are tested, what is the chance no defectives will appear? What is the chance the test batch will be forty per cent defective?

§ 20. *Conditional Probabilities*

The examples treated in the last section are stated with great finality. Sometimes, however, the statement of a question contains "provisos" which very materially alter the probability desired. For instance, the question, What is the

probability that Christmas falls on Monday? has been found to have the answer $\frac{1}{7}$. But the question, What is the probability that Christmas falls on Monday if it does not fall on either Friday or Saturday? is also a proper subject for the theory of probability, and it is immediately obvious that the answer is no longer $\frac{1}{7}$.

The answer to a question of this kind is called a "conditional [1] probability," and will be represented by the symbol $P_A(B)$, which may be read "the probability that B happens if A does." Problems of this sort are sometimes exceedingly difficult to solve, but in many cases they are almost as simple as if the restrictive conditions had not been applied. This is true, for instance, in the case of the question asked above. If Christmas does not fall on either Friday or Saturday it must fall on one of the remaining days of the week, of which there are five. As these are equally likely, the answer to the problem is $\frac{1}{5}$.

Another example is the following:

EXAMPLE 12.—*If, in the experiment in psychic research described in Example 11, the first card to appear is black but is called red by the medium, what is the chance that at the end of the trial there will be exactly three red cards in the red positions?*

The medium having called the first card red is left with three reds and four blacks. These she will call in some order, which is again our "standard order."

As the first card dealt was black, the only way in which the remainder can match up in exactly three reds is for all of the remaining red positions in the standard order to be matched with red cards. If this is done there are left three blacks and one red with which to fill the four black positions. It is obvious that the red card can be placed in any one of the four positions, after which the placing of the three black cards gives no new arrangement. Thus it is seen that the total number of permutations in the subgroup which satisfies the

[1] Sometimes "contingent."

conditions of the problem is 4. As the complete group consists of $P^{3,4}_{3,4} = 35$, the desired probability is $\frac{4}{35}$.

In other words, the medium's chance of scoring 75 per cent is reduced by half if her first attempt is wrong.

It is interesting to see how much greater her chances are if she gets the first card right. Therefore the following example may be considered:

EXAMPLE 13.—*If, in the experiment in psychic research described in Example 11, the medium places the first card correctly, what is her chance of having exactly three correct cards in each position?*

Note that the statement of this example is somewhat different from the statement of the preceding one. The preceding example stated definitely that the first card in the standard order was black, but was called red by the medium. As the problem is stated at present the color of the first card is not known. It is only known that the medium called it correctly. For the purposes of the problem, however, this makes no difference, because of the fact that the two colors are equally numerous and equally likely. The apparent difficulty can be overcome by speaking, not of " reds " and " blacks," but of " cards of the first color " and " cards of the second color," meaning by cards of the first color, cards of that color which first appears in the standard order.

The medium having first correctly called a card of the first color is left with three of the first color and four of the second, which she will call in a " standard order." If the cards appear in such a way as to misplace one card only of each color, a card of the second color must of necessity appear in some position which, in the standard order, is occupied by a card of the first color. As there are only three such positions remaining this can happen in just three ways. The remaining two positions of the first color in the standard order must then be filled by cards of the first color, after which there are left one card of the first color and three of the second with which to fill the four positions that are occupied by cards of the second color in the standard order. As the card of the first color may

be placed in either of the four positions, the total number of permutations favorable to the conditions of the problem is $3 \cdot 4 = 12$.

As for the total number of permutations in the complete group, this is easily seen to be $P^{3,\,4}_{3,\,4} = 35$. The answer to the problem is therefore $\frac{12}{35}$.

If, therefore, the medium guesses correctly at the first attempt, her chances of scoring exactly 75 per cent are increased by half.

If the problem had been stated explicitly, as was done in the case of Example 12, the argument would have been just the same except that the first color and second color would have been either red and black, respectively, or black and red, respectively, according to the exact conditions laid down. As has been said above, the reason why it is possible to treat the case generally is that the two colors are equally likely and equally numerous. A further example can be added in which this is not true:

EXAMPLE 14.—*If, in the psychic research experiment explained in Example 11, there are six red and two black cards, and if the first card to appear is red, but is called black by the medium, what is the chance that the black pile will contain only one red card?*

The standard order is now introduced by a black card. The card that first appears, however, is red. There are therefore left just $P^{6,\,1}_{6,\,1} = 7$ possible permutations, all of which are equally likely but not all of which match the standard order to the degree which is required by the conditions of the problem. They constitute the complete group required by the definition of "probability."

After the first card has been disposed of, the standard order contains five blacks and two reds. In order to satisfy the conditions of the problem the one remaining red card must appear in one of the two red positions, after which the positions of the black cards are immaterial. There are therefore just two permutations which satisfy the conditions of the problem. They constitute the group the probability of which is desired.

Before passing on to the next subject it should be noted that the method of solution used in these problems consists solely of the application of the definition of probability and the second general law of composition of events. Many problems of a much more complicated nature are capable of formal solution by the same means, though the actual numerical computation is often much more difficult. Success in treating problems of this sort is contingent on two essentials: a vivid and accurate mental picture of what is desired, and constant caution that the conditions regarding equal likelihood, mutual exclusiveness and independence are not violated.

Finally, it must not be thought that there is only one way of obtaining the solution of such problems as these. That is rarely true in any mathematical work. In the present instance the answers to the different problems are inter-related in such a way that some of them can be obtained from others by simpler processes than those which we have here employed. A little later on, when the ideas which are needed for this purpose have been introduced, the problem will again be considered in order to show how this can be done.

PROBLEMS

1. In Problem 2, § 19, if the connector has not come to rest within 0.66 second, what is the chance that it will come to rest within the next 0.33 second?

2. In Problem 2, § 19, if the connector has not come to rest within 0.66 second after starting, what is the chance that it will come to rest within the following intervals after starting:

1.00–1.50,
0.66–10.00,
0.20–0.50?

3. If, in the problem of Example 11, the first two cards drawn are black, but the medium calls one black and the other red, what is her chance of having a score of 75 per cent?

4. If, in the psychic research experiment, six red and two black cards are used, and if the first card to appear is black, but is called

red by the medium, what is the chance that she will score 75 per cent? Is the answer the same as in Example 14? Why?

5. If, in the psychic research problem, six red and two black cards are used, and if the first card to appear is red and is correctly called, what is the chance of a 75 per cent score?

6. The product of a lamp manufacturing concern has averaged 5 per cent bad over an entire year. A small retailer gets a shipment of twenty cartons of five. It does not follow, of course, that he has received just five bad lamps, for his "sample" of the gross product may have been better or worse than the average. A customer purchases one carton. How much greater is his chance of having all good lamps if the shipment was 1 per cent better than the average, than it would have been if the shipment had been 1 per cent worse than the average?

7. What is the chance of throwing an ace with an unsymmetrical die, if the ace is 5 per cent more likely than the adjacent sides, and the six 10 per cent less likely?

§ 21. *Compound Probabilities*

So far no cases have been considered which involved the simultaneous occurrence of more than one event. This will be the next subject of study. The general law which governs such cases reads as follows:

The probability that event A occurs and is accompanied[1] *by event B is the product of the probability that A occurs by the conditional probability that if A occurs B likewise occurs. In symbolic form*

$$P(AB) = P(A)P_A(B). \qquad (20)$$

[1] If A and B can occur simultaneously the law is true as stated. Sometimes there is a sequence of events as in the statement "a man is shot and dies." In this case "accompanied" must be replaced by "followed." When this is true the number given by this theorem represents the probability of A happening and being followed by B. The order cannot be reversed.

For instance, there is a certain chance that a man will be shot. *If he is shot* there is a chance that he will die. The product of these is the probability that he will die from a gunshot wound. It is not the same as the chance that he "dies and is then shot" — the latter being a much less common occurrence.

I would use the word "followed" except that it conveys a time-signification which is quite foreign to the subject.

This law is analogous to the second general law of composition of events and is capable of general proof as we shall see in § 47. For the present, however, we do not have the necessary equipment for such a general proof, and shall therefore content ourselves with a special case.

Let us begin with the consideration of a particular example exactly similar to Example 10:

EXAMPLE 15.—*The letters of the word* tooth *are written on cards, which are then thoroughly shuffled. If two are drawn in order, what is the chance that they yield the word* to*?*

The method of § 19 gives at once the answer $\frac{1}{5}$, but the argument can be phrased in a slightly different form. To get **to** it is necessary to get, first **t**, then **o**. But the first draw results in one of five cards, and the second draw in one of the four which remain. The various possibilities may then be listed schematically thus:

$$\frac{t^*}{o^*\ o^*\ t\ h} \qquad \frac{o}{t\ o\ t\ h} \qquad \frac{o}{t\ o\ t\ h} \qquad \frac{t^*}{t\ o^*\ o^*\ h} \qquad \frac{h}{t\ o\ o\ t}$$

In all, there are $5 \cdot 4$ of them. Of these possibilities the starred ones lead to **to**. There are $2 \cdot 2$ of them. Hence the result is $\frac{2 \cdot 2}{5 \cdot 4}$. But this fraction naturally falls apart into $(\frac{2}{5})(\frac{2}{4})$, the first term of which corresponds—not only in value but in the way in which it arrived as well—to the unconditional probability of drawing a **t**; while the second term likewise corresponds to the conditional probability of drawing an **o** *if* **t** *was already drawn*. Hence the example checks with the law.

In general, if the "event A" is itself a subgroup of n events which forms part of a complete group of m; and if every member of the complete group has associated with it a complete group of m' subsequent events, so that there are m of these "subsequent groups"; and finally if in *each* of these subsequent groups which are associated with "event A" there are n' which produce "event B"—if all these statements

are true the aggregate constitutes a total of mm' compound events of which a total of nn' produce "A followed by B." The desired probability is therefore $nn'/mm' = (n/m)(n'/m')$, of which the two factors are, by definition, " the unconditional probability of A " and " the conditional probability of B."

The theorem is of fundamental importance, and it must be repeated that our proof places restrictions upon it to which it is not, in fact, subject. For instance, if Example 15 read

EXAMPLE 16.—"*The letters* t, o, o, t, h *are written on cards, which are shuffled. One is drawn and the letter noted. If it is not a* t, *the* t's *are sorted out and discarded. Then a second shuffle and draw are carried out. What is the probability of* t, o*?*"

the possibilities would be

$$\frac{\text{t}^*}{\text{o}^*\ \text{o}^*\ \text{t}\ \text{h}} \qquad \frac{\text{o}}{\text{o}\ \text{h}} \qquad \frac{\text{o}}{\text{o}\ \text{h}} \qquad \frac{\text{t}^*}{\text{t}\ \text{o}^*\ \text{o}^*\ \text{h}} \qquad \frac{\text{h}}{\text{o}\ \text{o}}.$$

The events of the first group are no longer *all* associated with equally numerous subsequent groups: but though this violates the conditions of the " proof," it obviously does not invalidate the theorem. In § 22 we shall be able to extend the proof to cover this case, though it will still not be perfectly general.

Finally, it should be pointed out that the theorem can be extended to any number of consecutive events: the chance that *all* will occur in the prescribed order is the product of the proper set of probabilities. Some examples are given:

EXAMPLE 17.—*The letters of the word* tailor *are written on cards. The cards being first thoroughly shuffled, four are drawn in order. What is the probability that the result is* oral*?*

Of the six equally likely choices for the first letter only one is an **o**. Hence the chance of getting a group beginning with **o** is $\frac{1}{6}$. After this card has been drawn five remain. Hence if **o** is first drawn, the chance of drawing **r** next is $\frac{1}{5}$. Hence the chance of drawing **o** followed by **r** is $\frac{1}{30}$. If **o** and **r** are first drawn the chance of next drawing **a** is $\frac{1}{4}$. Therefore the chance of drawing **o r a** is $\frac{1}{120}$. Finally, if these three letters have appeared, the chance that the next trial will produce

1 is $\frac{1}{3}$. Thus the answer to the problem is $\frac{1}{360}$. Naturally this is the same result as was obtained in Example 9.

The solution of the problem as it is here given has required the use of the unconditional probability of drawing an **o**, together with three additional probabilities, each of which stands in the relationship of a conditional probability to all those which precede it.

EXAMPLE 18.—*The letters of the word* pepper *are written on cards. The cards having been thoroughly shuffled, four are drawn in order. What is the chance that the result is* peep *?*

In this case the chance of first drawing a **p** is $\frac{3}{6}$. Thereafter the chance of drawing an **e** is $\frac{2}{5}$, and the chance of drawing another **e** $\frac{1}{4}$. If all these appear the chance of drawing a **p** on the fourth attempt is $\frac{2}{3}$. Thus the result sought is $\frac{3}{6} \cdot \frac{2}{5} \cdot \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{30}$.

EXAMPLE 19.—*In the psychic research experiment of Example 11, what is the chance that the first card is correctly called, and that only one card of each color is incorrectly placed?*

The chance that the first card is correctly called is $\frac{1}{2}$. If so, the conditional probability that three cards in each pile are correct is found from Example 13 to be $\frac{12}{35}$. Therefore the answer to the problem is $\frac{1}{2} \cdot \frac{12}{35} = \frac{6}{35}$.

EXAMPLE 20.—*In the psychic research experiment of Example 11, what is the chance that the first card drawn is wrongly placed, and that all but two are correctly placed?*

The chance that the first card is wrongly placed is evidently $\frac{1}{2}$. If the first card was wrongly placed, the chance of having only two wrong is, by Example 12, $\frac{4}{35}$. Hence the solution of this problem is $\frac{1}{2} \cdot \frac{4}{35} = \frac{2}{35}$.

EXAMPLE 21.—*Each of two groups of cards contains four reds and four blacks. One group, which we shall call Group I, is first thoroughly shuffled, and laid out face down. Group II is then shuffled and its cards dealt out on top of those of Group I. What is the chance that the first card dealt from each group is black, and that of the remaining pairs just five are matched colors?*

The chance that the first card of Group I is black is $\frac{1}{2}$. If so, the chance that the card placed on it is black is $\frac{1}{2}$. Therefore, the chance that the first card is black and matched by a black is $\frac{1}{4}$. But if the first card is matched, the chance of having just two unmatched pairs is $\frac{12}{35}$, as is easily seen by comparison with Example 13. Therefore the answer to the problem is $\frac{1}{4} \cdot \frac{12}{35} = \frac{3}{35}$.

PROBLEMS

Using the time values of Problem 2, § 19, answer the following questions:

1. If the observer notes the operating times for 10 calls, what is the probability that all are less than 0.66 second?

2. What is the probability that just 2 of the calls have less than 0.66 second operating time?

3. Referring to Example 17, if each card drawn is replaced and shuffled before another is drawn, what is the probability of oral. Has it been increased or decreased by replacing the cards?

4. In Example 17 what is the chance of a result till if the cards are not replaced? If they are? Has replacing the cards increased or decreased the result?

The next three problems refer to the experiment explained in Example 21, except that it is assumed that six red and two black cards are used:

5. What is the probability that the first card of Group I is black, the first from Group II red, and that there are just two unmatched pairs?

6. What is the probability that the first cards are both red and that there are just two unmatched pairs?

7. What is the probability that the first cards are red and black respectively, and that there are just two unmatched pairs?

8. If in Example 21 the cards of Group I *are not* shuffled, but are laid out in order, reds being first; and if the cards of Group II *are* shuffled, what is the chance of just six matched pairs?

9. Under the conditions of Problem 8, what is the chance that the first pair are both red, and two of the remaining pairs unmatched?

10. Under the conditions of Problem 8, what is the chance that the first pair are both black, and two of the remaining pairs unmatched?

§ 22. *Alternative Compound Probabilities*

Many problems, which resemble those of the last section in form, require the use of Convention II in their solution.

EXAMPLE 22.—*The numbers* 1, 2, 3, 4, 5 *are written on cards, of which two are drawn without replacement. What is the chance that the combination thus drawn is even?*

Obviously, what the first card is does not matter, and the second is as likely to be one as another; hence the answer must be $\frac{2}{5}$. But suppose the problem is attempted by the argument of § 21. The possible results appear schematically as follows:

1	2	3	4	5
2* 3 4* 5	1 3 4* 5	1 2* 4* 5	1 2* 3 5	1 2* 3 4*

Obviously *any* number is admissible for the first choice, hence the unconditional probability of drawing an allowable first number is 1. But what is the conditional probability of a suitable second choice? Every first choice is associated with four possible second choices, of which *sometimes one and sometimes two are suitable.* Obviously the situation is not covered by the proof given for the fundamental theorem on compound probabilities.

On the other hand, it is easy to find the probability of "an even number beginning with **1**." It is $\frac{1}{5}\cdot\frac{2}{4} = \frac{1}{10}$. The same is true of "an even number beginning with **3**," or "with **5**"; while the probabilities for numbers beginning with **2** and **4** are each $\frac{1}{5}\cdot\frac{1}{4} = \frac{1}{20}$.

These things are mutually exclusive: therefore the chance of one or the other of the five happening is the sum of their separate probabilities, which is $\frac{2}{5}$, and as an even number can result in no other way, this must be the desired probability. Naturally, it checks the result obtained directly. In general:

If an event can be expressed as the sum of a number of alternative compound events, which compound events are mutually exclusive, and for which the separate probabilities can be found, the probability of the original event can be found by Convention II. It is essential that no manner of occurrence of the original event shall be overlooked.

Now any such process of division of the event A must be made by introducing extraneous conditions not contemplated in the original question. For instance, Example 22 neither expresses nor implies any condition upon the first number drawn. That is evident from the way the solution was obtained in the first sentence after the statement of the example. Yet each of the separate probabilities used in the second method of solution is obtained by introducing one of these "extraneous" events. If we call the result of the first drawing A, and "an even two-place number" B, the process appears symbolically as

$$P(B) = \sum_A P(A)\, P_A(B). \qquad (21)$$

It is particularly evident, in this symbolic form, that the events A, like the differential elements in an integral, are a sort of "catalytic agent," introduced for the purpose of enabling our computation to be carried out, though not themselves a part of the result. It should also be noted that, though all the probabilities $P_A(B)$ on the right-hand side are "conditional," the $P(B)$ to which they give rise is not.[1]

As for the range of summation implied in the Σ, it is always allowable to have it cover the complete group of events A; but if some $P_A(B)$ is zero, the A to which it corresponds may be omitted without error. Thus, if Example 22 had asked for the probability of an "even number less than 30," combinations beginning with 3, 4 or 5 could have been omitted;

[1] That is, not conditional in any way upon the set A: the whole process may be predicated upon the desire to obtain a probability which is conditional with respect to some event or events not entering the present discussion.

they could also equally well have been included since the accompanying conditional probabilities would have been zero.

Finally, a particular case which is of sufficient importance to merit special mention is that in which the extraneous group consists of an event A and its complement $\bar{A}$. Then the formula reads:

$$P(B) = P(A)\, P_A(B) + P(\bar{A})\, P_{\bar{A}}(B).$$

Thus in Example 22 the first number must be either *even* (A) or *not even* ($\bar{A}$), the respective probabilities being

$$P(A) = \tfrac{2}{5},$$
$$P(\bar{A}) = \tfrac{3}{5}.$$

As the conditional probabilities associated with them have already been said to be

$$P_A(B) = \tfrac{1}{4},$$
$$P_{\bar{A}}(B) = \tfrac{2}{4},$$

the solution of the example is

$$P(B) = \tfrac{2}{5}\cdot\tfrac{1}{4} + \tfrac{3}{5}\cdot\tfrac{2}{4} = \tfrac{2}{5},$$

as before.

Example 23.—*If the two groups of cards to which reference is made in Example 21 contain six red and two black cards each, what is the chance that the first pair is matched and the score is 75 per cent?*

In order that the conditions of this example may be fulfilled, the first card must be either red in both groups, or black in both groups. However, the chance that it is red in Group I is $\frac{3}{4}$, and if so, the chance that it is red in Group II is $\frac{3}{4}$ also. Thus the chance that it is red in both groups is $\frac{9}{16}$. On the other hand, the chance that it is black in Group I is $\frac{1}{4}$, and if so, the chance that it is black in Group II also is $\frac{1}{4}$, so that the chance of it being black in both groups is $\frac{1}{16}$.

These two events are mutually exclusive, and to that extent satisfy the conditions laid down upon the set of " events A " in (21). They do not constitute a complete set: it is possible for the first card of Group I to be red, and the first of Group II black, or *vice versa*; but the conditional probability of event B (that is, of " the first pair matched, and the score

WHEATON COLLEGE LIBRARY
WHEATON, ILLINOIS

75 per cent") is then zero. Hence it is not necessary to evaluate the probabilities of these remaining possibilities.

As for the conditional probabilities of 75 per cent matching in the two cases which are not trivial — that is, the quantities represented by $P_A(B)$ in (21) — these are easily found to be $\frac{10}{21}$ and $\frac{6}{7}$, respectively. Hence the answer to the problem is $\frac{9}{16}\cdot\frac{10}{21}+\frac{1}{16}\cdot\frac{6}{7}=\frac{9}{28}$.

EXAMPLE 24.—*If the two groups of cards to which reference is made in Example 21 contain six red and two black cards each, what is the chance that the first pair fails to match and the score is 75 per cent?*

The chance that the first card is red in Group I is $\frac{3}{4}$, and if so, the chance that it is black in Group II is $\frac{1}{4}$. The combination is one of two in which the cards of the first pair fail to match. Its probability is evidently $\frac{3}{16}$. The chance that the first card of Group I is black and that of Group II red is likewise $\frac{3}{16}$. These are the unconditional probabilities $P(A)$.

As for the conditional probabilities, the first is obviously identical with the solution of Example 14.[1] It is $\frac{2}{7}$. The

[1] As a matter of fact, so far as mathematical ideas are concerned, there is no difference whatever between the "standard order" which the medium sets up in the psychic research example, and the order in which the cards of Group I appear in Example 21 and those which follow it. But there is a psychological difference. For, if I am not mistaken, the medium who was confronted with the necessity of calling "red" six times and "black" but twice would be almost certain to say "red" the first time; so that we are not justified in saying that any one of her eight *words* is as likely as any other to be first, though we are justified in the assertion that any one of the eight cards of Group I is as likely as any other to be first.

Perhaps this fact may be used to emphasize the point about the axiom of equal likelihood to which we have already referred in § 3. There is a vast difference between the assertions "Two events are equally likely when they differ in no known pertinent attribute" and "Two events are equally likely when we do not know what difference their pertinent attributes make." The first statement leaves us with many situations to which the fundamental method of measuring probability cannot be applied—as in the present instance where, though I feel certain the psychological bias exists, I am wholly unable to state its extent. The second virtually says, "Whenever it is impossible to measure the probabilities of a group of events, they are equally likely," which is a sheer absurdity.

It is the failure, on the one hand, to clearly *express* this idea, and on the other to clearly *grasp* it, which has led to much of the argument over the dogmas of "insufficient reason" and "cogent reason" to which reference will again be made in § 48.

other is also found to be $\frac{2}{7}$. The substitution of these values in (21) leads to the result $\frac{3}{16}\cdot\frac{2}{7}+\frac{3}{16}\cdot\frac{2}{7}=\frac{3}{28}$.

§ 23. *Some Instructive Illustrations; The Psychic Research Problem*

Certain interesting relations exist between the probabilities which have been obtained in the case of the psychic research example. In the first place, confining our attention to the case of four red and four black cards, there is obviously no pertinent distinction between the four red positions in the standard order which would make a black card more likely to fall in one of them than another. Hence the unconditional probability that the medium is wrong the first time she says "red" must be the same as the unconditional probability that she is wrong the second time she says "red," or the third, or the fourth. Furthermore, since the card that falls in either of these positions is just as likely to be black as red, this unconditional probability is $\frac{1}{2}$. All this is axiomatic.

Likewise there is no distinction between the first occurrence of the word "red" and its subsequent occurrences which would cause the conditional probability of a 75 per cent score to be different if one were known to have been incorrectly called rather than another. That is, if by accident we happened to observe that a card which the medium called "red" was actually black, it would not matter whether it was the *first*, or some subsequent occurrence of the word: the probability of a 75 per cent score would in either case be $\frac{4}{35}$, as obtained in Example 12.

But if a 75 per cent score is to be obtained at all, one or the other of the red positions in the standard order must be filled by black cards. As the four *compound* events ("first red position occupied by a black card and a 75 per cent score," "second red position occupied by a black card and a 75 per cent score," etc.) are mutually exclusive, it follows that the *unconditional* probability of a 75 per cent score is the sum of four compound probabilities, each equal to $\frac{1}{2}\cdot\frac{4}{35}$. This

works out to be $\frac{8}{35}$, which is, of course, the same as the result already obtained in Example 11.

Another result which can be obtained out of this same argument is the following: To say that the first card is *correctly* called and the score is 75 per cent, is equivalent to saying that either the second, the third, or the fourth black card is *incorrectly* called, the rest being correct in each case. The probability of each of these alternatives, however, is $\frac{1}{2} \cdot \frac{4}{35} = \frac{2}{35}$. Hence the probability that the first card is correctly called and the score is 75 per cent works out to be $\frac{2}{35} + \frac{2}{35} + \frac{2}{35} = \frac{6}{35}$; a result which has already been obtained.

Another result may easily be obtained from the use of (20), if we regard "event A" as meaning that the first card is correctly called, and "event B" that a 75 per cent score is obtained. In this case, (20) states that the probability that the first card is called correctly and the score is 75 per cent is the product of the unconditional probability of calling the first card correctly — which is known to be $\frac{1}{2}$ — by the conditional probability of obtaining such a score under these circumstances, which we denote merely by $P_A(B)$. Equating $\frac{1}{2}P_A(B)$ to the result obtained in the last preceding paragraph, it is found that $P_A(B) = \frac{12}{35}$. This again is identical with the result of Example 13.

Many more relationships of this sort can be built up between the numbers already obtained. These are sufficient, however, to illustrate to what extent the theorems developed in the last few sections are capable of simplifying the solution of problems of this sort.

§ 24. *Some Instructive Illustrations; A Generalization of the Psychic Research Problem*

After having used the psychic research experiment explained in Example 11 so profusely for illustrative purposes, it would be unnatural to pass it by finally without obtaining a solution of somewhat greater generality than the special cases already considered. Hence the following general case is given:

EXAMPLE 25.—*Each of two sequences contains m events of one kind and n of another. The sequences are placed in one-to-one correspondence by a method of choice which is not influenced by the characteristics that differentiate the two kinds. What is the chance that there just p pairs which fail to match?*

Phrased in terms of "cards" and "medium" it reads:

EXAMPLE 26.—*If, in the psychic research experiment explained in Example 11, there are m red and n black cards, what is the chance of just p incorrect cards in each pile?*

The simplest way to solve the problem is by means of the fundamental definition of probability—just as Example 11 itself was solved. The order in which the medium calls the cards forms a "standard order," while the order in which the cards actually appear may be any one of the $P^{m,n}_{m,n}$ possible permutations of m red and n black things. These represent a complete group of equally likely and mutually exclusive events. Certain of these permutations fail to match the standard order in exactly p of the m reds and exactly p of the n blacks, and thus form a subgroup which meets the conditions of the problem. Put in other words this says that those positions which are red in the standard order must be filled by some possible permutation of p blacks and $m - p$ reds, while those which are black in the standard order must be filled by some possible permutation of p reds and $n - p$ blacks. Each of the possible ways in which the standard red positions can be filled is capable of association with each of the possible ways in which the standard black positions can be filled. Therefore the total number of events in the subgroup is

$$P^{m-p,p}_{m-p,p} \cdot P^{n-p,p}_{n-p,p}.$$

The desired probability is therefore

$$P(p) = \frac{P^{m-p,p}_{m-p,p} \cdot P^{n-p,p}_{n-p,p}}{P^{m,n}_{m,n}} = \left(\frac{m!n!}{p!}\right)^2 \frac{1}{(m-p)!(n-p)!(m+n)!}.$$

This can also be written in the form

$$P(p) = \frac{C_p^m C_p^n}{C_m^{m+n}}, \tag{22}$$

a form which is itself capable of logical interpretation.

One of the many elegant relationships between binomial coefficients can be obtained from (22). In a trial of the sort under consideration there must either be no incorrect cards in each pile, or one incorrect card in each pile, or two or three or some other number not exceeding the smaller of the integers m and n. This means that the sum

$$\sum_{p=0}^{m \text{ or } n} P(p)$$

is equal to unity, or

$$\sum_{p=0}^{m \text{ or } n} \frac{C_p^m C_p^n}{C_m^{m+n}} = 1.$$

This equation consists on the left-hand side of a number of fractions, and on the right-hand side of the integer 1. The validity of the equation will remain unchanged if every term in both members is multiplied by *the same* factor. Suppose C_n^{m+n} is chosen as this factor: then the equation becomes

$$\sum_{p=0}^{m \text{ or } n} C_p^m C_p^n = C_n^{m+n}.$$

This equation says that if any two columns of Appendix III are chosen and corresponding entries multiplied by one another until the end of the shorter column is reached, the sum of all the partial products thus obtained will itself be a binomial coefficient. Moreover, it must be in the particular column in Appendix III, the heading of which is the sum of the headings of the two columns chosen, and in a row which is denoted by the same number as one of the column headings.

For example, if the columns headed 3 and 8 are chosen, the entries (as far down as the end of the shorter column) and their partial products are

3	8	Product
1	1	1
3	8	24
3	28	84
1	56	56
		Sum = 165

The sum of the two column headings being $3 + 8 = 11$, the law expressed by (22) says that the sum of the partial products, 165, is equal to C_3^{11} or C_8^{11}. By reference to Appendix III, it will be found that each of these numbers is actually 165.

In stating the law the upper index on the sign of summation has been written as "m or n," meaning thereby the smaller of these two integers. Suppose for the moment that the smaller integer is m. Then as soon as p exceeds m, the binomial coefficient C_p^m vanishes. Hence, if the number of terms were extended beyond the limit fixed by the smaller integer, the result would be merely to add on a certain number of zeros. This, of course, would not in any way affect the sum. In other words, the upper limit of summation can be taken as anything at all, provided it be not less than the smaller of m and n. Thus

$$\sum_{p=0}^{m} C_p^m \; C_p^n = C_n^{m+n},$$

$$\sum_{p=0}^{n} C_p^m \; C_p^n = C_n^{m+n},$$

$$\sum_{p=0}^{\infty} C_p^m \; C_p^n = C_n^{m+n},$$

are all equally valid.

A word of caution should be added about one step in the process by which this result has been obtained. The transfer of C_n^{m+n} from the left-hand side of the equation to the right-hand side was accomplished by multiplying both sides by this quantity. In doing this the fact was stressed that each individual term on the left-hand side must of necessity be multiplied by *the same* number, otherwise the result would be incorrect. This means, so far as the shorthand notation is concerned, that the general term

$$\frac{C_p^m \; C_p^n}{C_n^{m+n}}$$

can be multiplied by anything whatever which does not vary with p; that is, which does not change from term to term. The factor in the denominator does not involve p and therefore could be taken outside the sign of summation by this process. But if the general term had been, for example,

$$\frac{C_p^m \; C_p^n}{C_p^{m+n}},$$

it would have been quite improper to multiply each term by C_p^{m+n} and arrive at the result

$$\sum_{p=0}^{m \text{ or } n} C_p^m \, C_p^n = C_p^{m+n}.$$

In fact, this last equation not only is not correct, but it cannot even be interpreted, a fact concerning which the student can easily satisfy himself by attempting to assign numerical values to the letters.

§ 25. *Some Instructive Illustrations; The Problem of Independent Trials*

A very fundamental class of compound events is that in which the chance of an event occurring is not in any way influenced by what has already occurred: that is, where every event is quite independent of the rest. Dice problems are of interest in the subject of probability largely because they typify this class of events. The following is a simple example:

EXAMPLE 27.—*What is the chance of throwing an ace exactly once in six throws of a die?*

This problem can be solved by the use of alternative compound probabilities. The possible cases are

$$+ - - - - -, \quad - + - - - -, \quad - - + - - -,$$
$$- - - + - -, \quad - - - - + -, \quad - - - - - +.$$

Taking the first case, the probability of throwing an ace on the first throw is $\frac{1}{6}$. The probability of not throwing an ace on any of the remaining throws is $(\frac{5}{6})^5$, since the events are all independent. The probability of the first compound event is therefore $\frac{1}{6}(\frac{5}{6})^5$.

In the second case the answer is obtained in the form $\frac{5}{6}(\frac{1}{6})(\frac{5}{6})^4$, which is exactly the same as before except for the order in which the fractions appear.

The remaining alternative cases can be similarly treated, and in each instance the answer comes out the same. Since these cases are all mutually exclusive the chance that one or

the other occurs—which is what the problem asks for—is $6 \cdot \frac{1}{6} \cdot (\frac{5}{6})^5 = (\frac{5}{6})^5$.

Let us see what this problem really teaches. Since the throws are independent the chance of throwing one ace and five "not aces" in some preassigned order is always $\frac{1}{6}(\frac{5}{6})^5$, regardless of the order chosen. Hence it is necessary to add together as many of these equal quantities as there are distinct orders in which the result may appear. The number of distinct orders, however, is just the number of permutations of six things of which five are alike [1] and one different, that is $P^{5,1}_{5,1}$.

In general, if the probability of throwing exactly n aces in m throws had been asked for, there would be as many possible orders in which the result might occur as there are ways of permuting n things of one kind and $m - n$ of another, that is, $m!/n!(m-n)!$. This happens to be numerically equal to C^m_n and therefore can be conveniently represented by that symbol. The solution of the general problem therefore consists of the sum of C^m_n terms, each corresponding to the probability of throwing m aces and $m - n$ "not aces" in some preassigned order. Since the throws are independent each of these terms is the product of n equal factors $\frac{1}{6}$ and $m - n$ other equal factors $\frac{5}{6}$; that is, to $(\frac{1}{6})^n(\frac{5}{6})^{m-n}$. The answer is therefore

$$C^m_n \, (\tfrac{1}{6})^n \, (\tfrac{5}{6})^{m-n}.$$

It is now a simple matter to extend this formula into the following general theorem:

If the probability of an event occurring in a single trial is p, the chance that it occurs exactly n times in m INDEPENDENT *trials is*

$$P_m(n) = C^m_n \, p^n \, (1 - p)^{m-n}. \tag{23}$$

This is one of the fundamental theorems of the Theory of Probability. It will receive full discussion later.

[1] That is, "not aces." Their distinguishing characteristics are non-essential.

§ 26. *Some Instructive Illustrations; A Generalization of the Problem of Independent Trials*

We can easily derive as a corollary to this theorem another which is frequently useful. Speaking again in terms of the dice problem which we have considered, but writing n_1 where we previously wrote n, we note that the $m - n_1$ "not-aces" are composed of deuces and other things. The chance that there are just n_2 deuces among them is just

$$C_{n_2}^{m-n_1} \left(\tfrac{1}{5}\right)^{n_2} \left(\tfrac{4}{5}\right)^{m-n_1-n_2},$$

since the $m - n_1$ trials are certainly independent and the probability of a deuce appearing is $\frac{1}{5}$ when ace is known not to appear. We must not overlook the fact that what we obtain in this manner is the conditional probability of n_2 deuces in m trials *if there are known to be just* n_1 *aces*. Hence we conclude that the chance of just n_1 aces and n_2 deuces is the product of this expression by the unconditional probability of n_1 aces.

If we go another step we find as the conditional probability of n_3 threes in m trials, *if there are known to be just* n_1 *aces and* n_2 *deuces*, is

$$C_{n_3}^{m-n_1-n_2} \left(\tfrac{1}{4}\right)^{n_3} \left(\tfrac{3}{4}\right)^{m-n_1-n_2-n_3};$$

and by multplying this factor in with the two already found we can obtain the probability of exactly n_1 aces, n_2 deuces and n_3 threes.

Carrying this process on step by step we eventually find the probability of exactly n_1 aces, n_2 deuces, . . . , n_6 sixes. After common factors have been cancelled out the result is

$$P_m(n_1, n_2, \ldots, n_6) = P_{n_1, n_2, \ldots, n_6}^{n_1, n_2, \ldots, n_6} \left(\tfrac{1}{6}\right)^{n_1} \left(\tfrac{1}{6}\right)^{n_2} \ldots \left(\tfrac{1}{6}\right)^{n_6},$$

it being understood, of course, that $n_1 + n_2 + \ldots + n_6 = m$.

This, too, is easily converted into a general theorem, though the fact that the six faces of the die are equally likely makes it rather difficult to guess what it is to be. For that reason, and also because the alternative form of proof is interesting

in itself, we sketch a proof of the general case by a different line of argument.

If we have a complete and mutually exclusive set of events the individual probabilities of which are $p_1, \ldots, p_s$, and if we make m independent trials, the chance that the first event occurs just n_1 times, the second n_2 times, and so on *in a specified order* is just

$$p_1^{n_1} p_2^{n_2} \ldots p_s^{n_s},$$

no matter what order may be specified. Hence to get the chance of the first event occurring n_1 times, the second n_2 times, and so on, *regardless of order*, it is only necessary to multiply this quantity by the number of possible permutations. This gives us immediately the theorem:

If the events denoted by the subscripts 1, 2, . . . , s *are mutually exclusive and form a complete set, and if their respective probabilities of occurrence are* $p_1, p_2, \ldots, p_s$, *the chance that they will occur with the frequencies* $n_1, n_2, \ldots, n_s$ *in* $m = n_1 + \ldots + n_s$ *independent trials is*

$$P_m(n_1, n_2, \ldots, n_s) = P^{n_1, n_2, \ldots, n_s}_{n_1, n_2, \ldots, n_s} p_1^{n_1} p_2^{n_2} \ldots p_s^{n_s}. \qquad (24)$$

If there are just two events this formula reduces to (23); while if there are six equally likely events it reduces to the result obtained in our consideration of the tossing of a die.

§ 27. *Some Instructive Illustrations; A Typical Urn Problem*

In dealing with the last example, considerable emphasis was placed upon the fact that the trials were independent of one another. That a different result is obtained if this condition is not satisfied may be illustrated by the following example:

Example 28.—*An urn contains five red and ten black balls. Eight of these are drawn out and placed in another urn. What is the chance that the latter then contains two red and six black balls?*

This example resembles the former one in that it might be very simply stated as, What is the chance of drawing

exactly two red balls in eight trials? It differs from the former in that the trials are not independent; that is, the chance of drawing a red ball on the first attempt is $\frac{5}{15}$, while the chance of drawing a red ball on the second attempt is either $\frac{4}{14}$ or $\frac{5}{14}$ according as the first ball was red or black.

The solution to the problem, however, is not hard to obtain. There are $P_{6,2}^{6,2} = C_2^8$ orders in which the two red and six black balls may appear. If each of these orders is separately considered the probability of drawing the balls in exactly this order may be found. The sum of all these terms will then be the desired answer. This appears to require the computation of $C_2^8 = 28$ terms—which would involve a considerable amount of labor. Fortunately, however, the consideration of a very few terms serves to show their law of formation and makes the complete solution quite simple.

Three of the possible 28 orders are obviously

r r b b b b b b,
b b b b b b r r,
b b r b r b b b.

Consider the first of these. The chance of choosing a red ball first is $\frac{5}{15}$. *If this is done*, the chance of choosing a red the next time is $\frac{4}{14}$. *If both these events have taken place*, the chances that the next six balls are each black are[1] $\frac{10}{13}$, $\frac{9}{12}$, $\frac{8}{11}$, $\frac{7}{10}$, $\frac{6}{9}$, $\frac{5}{8}$, respectively. Therefore the chance that all of these events take place is

$$\frac{5}{15} \cdot \frac{4}{14} \cdot \frac{10}{13} \cdot \frac{9}{12} \cdot \frac{8}{11} \cdot \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{5}{8}.$$

If the second case is considered the answer takes the form

$$\frac{10}{15} \cdot \frac{9}{14} \cdot \frac{8}{13} \cdot \frac{7}{12} \cdot \frac{6}{11} \cdot \frac{5}{10} \cdot \frac{5}{9} \cdot \frac{4}{8};$$

while if the third group is considered the result is

$$\frac{10}{15} \cdot \frac{9}{14} \cdot \frac{5}{13} \cdot \frac{8}{12} \cdot \frac{4}{11} \cdot \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{5}{8}.$$

Now the remarkable thing about these three expressions is, that although the separate fractions differ, the product is

[1] Note that these are all conditional probabilities, and are not equal one to another as in the case of independent trials.

the same for every case. As a matter of fact, this property is common to all the terms. Every time a ball is chosen, whether red or black, the number of balls remaining in the urn is reduced by one so that the denominator of the next conditional probability is also reduced by one. This means that no matter what the order of choice may be the array of factors in the denominator will always be the same. Likewise whenever a red ball is chosen the numerator for the next conditional probability of a red ball appearing is reduced by one. But the numerator of the next conditional probability for a black choice remains unchanged. Thus to the red choices correspond fractions the numerators of which are always 5 and 4, while to the black choices correspond numerators 10, 9, 8, 7, 6 and 5. As the permutation of these factors does not affect the magnitude of their product, it follows that every order of choice has the same probability as every other. In other words all orders are equally likely. The answer to the problem is therefore

$$C_2^8 \frac{5!\,10!\,7!}{3!\,4!\,15!} = \frac{140}{429}.$$

This same answer can also be obtained in another way. Suppose the balls are all tagged to establish their identity. There are then P_8^{15} ways in which eight can be drawn out, all of these ways being equally likely.[1] In order to solve the problem by means of the fundamental definition of probability it is only necessary to determine how many of these ways represent just two red balls and six black balls. This number may be found by observing that there are C_2^5 ways of choosing two red balls for the group and C_6^{10} ways of choosing six black balls. This makes a total of $C_2^5\,C_6^{10}$ different *combinations* of tagged balls which satisfy the condition of the problem.

[1] This may be taken, either as an intuitive fact, or as the result of computation; for if an order is specified, the chance of drawing in just that order is

$$\tfrac{1}{15}\cdot\tfrac{1}{14}\cdot\tfrac{1}{13}\cdots\tfrac{1}{8},$$

whatever the order may have been.

Each of these, however, is capable in itself of P_8^8 permutations. Thus, the total number of permutations satisfying the conditions of the problem is found to be $C_2^5 C_6^{10} P_8^8$, and the answer to the problem is obtained as

$$\frac{C_2^5 C_6^{10} P_8^8}{P_8^{15}}.$$

This is easily reduced to the form already found.

Just as in the preceding section, it is possible to generalize this solution and obtain a working theorem. Suppose the problem had read:

An urn contains m red and n black balls. If $p + q$ are drawn in order and placed in another urn, what is the chance that the latter contains just p red balls and q black ones?

If the balls were all tagged there would be P_{p+q}^{m+n} equally likely permutations. Any trial would give some one of these.

There would also be a total of $C_p^m C_q^n$ combinations of p red and q black balls each of which would be capable within itself of P_{p+q}^{p+q} permutations. Therefore the total number of permutations which satisfy the conditions of the problem and therefore compose the desired subgroup is $C_p^m C_q^n P_{p+q}^{p+q}$. Dividing the number of elements in the subgroup by the number of elements in the complete group it is found that the probability of drawing exactly p reds and q blacks is

$$P_{m,n}(p, q) = \frac{C_p^m C_q^n P_{p+q}^{p+q}}{P_{p+q}^{m+n}} = \frac{C_p^m C_q^n}{C_{p+q}^{m+n}}.$$

In words this theorem reads:

If a group of m things of one kind and n things of another exists, and if this group is reduced by eliminating one thing at a time, the thing being chosen quite without respect to those characteristics which differentiate kind from kind, the probability that the first $p + q$ stages will remove p things of the first kind and q of the second is

$$P_{m,n}(p, q) = \frac{C_p^m C_q^n}{C_{p+q}^{m+n}}. \tag{25}$$

A useful relation among the binomial coefficients can be obtained from (25) by an argument exactly analogous to that used in § 24. Suppose we denote the total number of withdrawals by r, so that we have the relationship $p + q = r$. Among the r things withdrawn there must be, either none of the first kind and r of the second, or one of the first and $r - 1$ of the second, or some other of the set of obvious possibilities. This leads us at once to the equation

$$\sum_{p=0}^{r} P_{m,n}(p, r-p) = 1;$$

or, upon substituting (25) in this equation and noting that the denominator is the same for every term,

$$\sum_{p=0}^{r} C_p^m C_{r-p}^n = C_r^{m+n}. \tag{26}$$

We shall find this formula of service in a later section.

§ 28. *Some Instructive Illustrations; Another Typical Urn Problem*

Suppose the example considered in the last section is modified to read:

EXAMPLE 29.—*An urn contains five red and ten black balls. Eight times in succession a ball is drawn out but it is replaced before the next drawing takes place. What is the probability that the balls drawn were red on two occasions and black on six?*

Since the balls are replaced before the next drawing takes place the condition of the urn is always the same just before every trial, and therefore the chance of drawing a red ball or a black ball is the same for each of the trials. In other words, the trials are completely independent. The theorem developed in § 24 therefore applies to this case.

The chance of drawing a red ball is $\frac{1}{3}$ and the chance of drawing a black ball $\frac{2}{3}$. Hence the chance of drawing exactly two reds and six blacks in eight trials is

$$C_2^8 (\tfrac{1}{3})^2 (\tfrac{2}{3})^6 = \tfrac{1792}{6561}.$$

This answer is considerably smaller than that obtained when the balls were not replaced. Decimally, the answer to the present problem is 0.273 while the answer to the other was 0.326.

§ 29. *Some Instructive Illustrations; A Problem in Matching*

A problem of a greater difficulty, which needs repeated use of alternative compound probabilities for its solution, is the following:

EXAMPLE 30.—*From a thoroughly shuffled pack of cards m are dealt and laid in a row face down in the order of appearance. This pack is then laid aside. Another pack is taken and m cards are dealt on top of the first m. In this way m pairs of cards are obtained. The cards in any pair may be either of like or of unlike color. What is the chance that there are exactly n matched pairs?*

This example bears an obvious resemblance to Examples 21 and 26, but differs from them in a very important respect which makes it much more difficult to solve. In the previous problems the number of red and black cards was both known and known to be equal in both sequences, whereas in the present case the number of reds and blacks dealt from the first pack is neither known nor known to be equal to the number dealt from the second pack.

As the number of reds dealt from the first pack is unknown, the natural thing to do is to assign a letter r to represent it. So far as is known r may have any value between zero and m.

Next we denote by $P(r)$ the chance of the assumed number of reds being the true one; and by $P_r(n)$ the chance of n matched pairs if it is. Then

$$P(n) = \sum_{r=0}^{m} P(r)\, P_r(n), \qquad (27)$$

$P(n)$ being the symbol chosen for the answer to the problem. This is a use of (21).

Now obviously $P(r)$ is a special case of formula (25), the

p and q of (25) being r and $m - r$, respectively, while both m and n are 26: that is,

$$P(r) = \frac{C_r^{26} C_{m-r}^{26}}{C_m^{52}}. \tag{28}$$

Hence $P(n)$ can be found if $P_r(n)$ can be found, so one of the indefinite features of the problem has been removed. The next step will eliminate another.

If there are exactly n matched pairs, there must be either n pairs of red cards and no black pairs, or $n - 1$ red pairs and one black pair, or some other combination of numbers having the sum n. If, then, $P_r(k, n - k)$ is the probability of just k red and $n - k$ black pairs, when there are r reds in the bottom row,

$$P_r(n) = P_r(0, n) + P_r(1, n - 1) + \ldots = \sum_{k=0}^{n} P_r(k, n - k). \tag{29}$$

The final step is to find $P_r(k, n - k)$. This is done by an argument exactly like that of § 22.

The order in which the bottom row of cards appears is unknown; but whatever that order is, it may be called the "standard order." On the other hand, the top row is some possible permutation of m cards chosen from the 52 cards of the deck. If the separate identities of all these cards are taken into account, the total number of permutations of this sort is $P_m^{52} = 52!/(52 - m)!$. These permutations are all equally likely, they are mutually exclusive, and they form a complete set. Therefore the desired probability $P_r(k, n - k)$ can be found by finding the subgroup which matches the standard order in exactly k reds and $n - k$ blacks. To find this number, it is easiest to find the number of possible *combinations*, and then the number of permutations of which each combination is capable within iself. The product is the desired number of *permutations*.

The k red cards which match reds in the standard order can be chosen in C_k^{26} ways, without taking account of order. With them may be combined any one of C_{r-k}^{26} combinations of blacks. Then, from the remaining $26 - r + k$ blacks the $n - k$ which match can be chosen in C_{n-k}^{26-r-k} ways; while the

$(m - r) - (n - k)$ reds which fall on blacks can be chosen in $C^{26-k}_{(m-r)-(n-k)}$ ways from among the $26 - k$ reds which remain. The total number of *combinations* of cards which may be dealt in the top row is therefore

$$C^{26}_{k}\, C^{26}_{r-k}\, C^{26-r+k}_{n-k}\, C^{26-k}_{(m-r)-(n-k)}.$$

Those cards which stand in the red positions of the standard order can be permuted among themselves in every possible way without in any way affecting the number of matches. The same is obviously also true of the cards which stand in the black positions of the standard order. The number of permutations of the first kind is $P_r = r!$ and the number of permutations of the second kind is $P_{m-r} = (m - r)!$. When the product of four C's written above is multiplied by these two P's it gives the desired subgroup. Hence:

The probability of matching the standard order in exactly k reds and $n - k$ blacks if the standard order contains exactly r reds and $m - r$ blacks is [1]*:*

$$P_r(k, n-k) = \frac{C^{26}_{k}\, C^{26}_{r-k}\, C^{26-r+k}_{n-k}\, C^{26-k}_{(m-r)-(n-k)}\, P_r\, P_{m-r}}{P_m^{52}}. \qquad (30)$$

[1] This formula has been worked out without the slightest reference to the limits within which the numbers k and $r - k$ must be confined. That there are such limits is obvious. For instance, the number of red pairs cannot be negative, which means that k must be greater than or equal to zero. Similarly, the number of black pairs cannot be negative, which means that $n - k \geqq 0$ or $k \leqq n$. On the other hand the number of red pairs cannot exceed the number of red cards in the standard order; that is, $k \leqq r$, and the number of black pairs cannot exceed the number of blacks in the standard order, which means that $n - k \leqq m - r$. If any one of these four conditions is violated $P_r(k, n - k)$ must be zero.

We have already noted some cases where formulæ of this sort automatically took the correct value zero when the sensible limits upon the variable quantity were transgressed. It is interesting to note that (30) actually vanishes when any of the above conditions is violated. For instance, for $k < 0$, C^{26}_{k} vanishes. If $k > r$, C^{26}_{r-k} vanishes. If $k > n$, C^{26-r+k}_{n-k} vanishes. If $(n - k) > (m - r)$, $C^{26-k}_{(m-r)(-(n-k)}$ vanishes.

If it were not for this fact, when we come to substitute the values of $P_r(k, n-k)$ in (29) it would be necessary to use formula (30) only for those values of k which lie within what we have called sensible limits. However, since (30) actually takes the correct value zero when these limits are transgressed, we may if we desire substitute it algebraically into (29) without worrying about this point. This fact simplifies the notation in equations (31), (32) and (33).

All that now remains is to collect and simplify the results. First, note that

$$\frac{P_r^r P_{m-r}^{m-r}}{P_m^{52}} = \frac{1}{C_r^m C_m^{52}}.$$

Then substituting this in (30), and (30) in (29), gives

$$P_r(n) = \sum_{k=0}^{n} \frac{C_k^{26} C_{r-k}^{26} C_{n-k}^{26-r+k} C_{m+k-n-r}^{26-k}}{C_r^m C_m^{52}}. \tag{31}$$

As the denominator does not vary with k, it may be taken outside the sign of summation, giving

$$P_r(n) = \frac{1}{C_r^m C_m^{52}} \sum_{k=0}^{n} C_k^{26} C_{r-k}^{26} C_{n-k}^{26-r+k} C_{m+k-n-r}^{26-k}. \tag{32}$$

Finally, substituting (32) and (28) in (27), and factoring out those terms which do not depend upon r, gives

$$P(n) = \frac{1}{(C_m^{52})^2} \sum_{r=0}^{m} \frac{C_r^{26} C_{m-r}^{26}}{C_r^m} \sum_{k=0}^{n} C_k^{26} C_{r-k}^{26} C_{n-k}^{26-r+k} C_{m+k-n-r}^{26-k}. \tag{33}$$

This is the desired answer.

§ 30. *Some Instructive Illustrations; An Example in Computation*

The formula (33) is quite complicated even when written in shorthand notation. To write it without the use of signs of summation would be next to impossible. Although the illustration itself has no great importance beyond its value as an example of a method of attack which must frequently be resorted to in complicated cases, it is probably wise to carry out the numerical computation of a special case in order, in the first place, to impress more clearly what the notation means, and in the second place, to show how it almost automatically builds up a scheme of numerical computation.

To this end we choose the following special case of Example 30:

EXAMPLE 31.—*If eight cards are dealt from each of two packs as explained in Example 30, what is the chance that there are exactly four matched pairs?*

The solution of this special case is, of course, given by substituting the values $m = 8$ and $n = 4$ in (33). There then remain two letters

r and k in the formula. According to the limits on the signs of summation, r is to take every value from 0 to 8 and k is to take every value from 0 to 4. In order to accommodate the numbers dependent upon these two variables it is necessary to have a table, the columns of which are headed 0, 1, 2, 3 and 4 to correspond to the values of k and the rows of which are numbered from 0 to 8 to correspond to the values of r. As a matter of fact more than one such table is required before the computation is complete.

TABLE III

COMPUTATION OF $C_k^{26} C_{r-k}^{26}$

r	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
0	1				
1	26	26			
2	325	676	325		
3	2,600	8,450	8,450	2,600	
4	14,950	67,600	105,625	67,600	14,950
5	65,780	388,700	845,000	845,000	388,700
6	230,230		4,858,750	6,760,000	4,858,750
7	657,800			38,870,000	38,870,000
8	1,562,275				223,502,500

The first of these tables (Table III) is devoted to the computation of the probability

$$C_k^{26} C_{r-k}^{26}.$$

When k is zero the product reduces to C_r^{26}; therefore the first column in the table is obtained by merely copying from Appendix III, taking account of the fact, of course, that C_0^{26} is 1. The second column is the product of the constant factor C_1^{26} by the variable factors C_{r-1}^{26} But C_1^{26} is the second entry in the first column, while the variable factors are themselves the successive entries in the same column, except that they are displaced one unit because the subscript is $r-1$ instead of r. This means that each integer in the first column is multiplied by 26, and the product recorded, not opposite the integer multiplied, but one space lower down. Similarly the third column is obtained by multiplying each element of the first column by 325 and recording the results two places below the row from which they were obtained. The other columns are computed in a similar manner. Only five entries have been placed in each column,

since what we have called the " sensible limits " upon k show that no other entries are needed.

The next step is the computation of C_{4-k}^{26-r+k}. This is carried out

TABLE IV

COMPUTATION OF C_{4-k}^{26-r+k}

r	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
0	14,950				
1	12,650	2,600			
2	10,626	2,300	325		
3	8,855	2,024	300	26	
4	7,315	1,771	276	25	1
5		1,540	253	24	1
6			231	23	1
7				22	1
8					1

in Table IV, and consists merely of entering in each column a number taken from Appendix III.

The third step is the computation of C_{4+k-r}^{26-k}. This again consists merely of writing down numbers taken from Appendix III. The

TABLE V

COMPUTATION OF C_{4+k-r}^{26-k}

r	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
0	14,950				
1	2,600	12,650	10,626		
2	325	2,300	10,626		
3	26	300	2,024	8,855	
4	1	25	276	1,771	7,315
5		1	24	253	1,540
6			1	23	231
7				1	22
8					1

result is given in Table V. It will be noted that the numbers omitted at the bottom of each column, like those omitted at the tops of the columns of Table III, are all zero. It is useless, therefore, to compute the factors by which these zeros would be multiplied in evaluating

(33). This fact accounts for the remaining empty spaces in the tables.

It is next necessary to multiply together the corresponding entries in these three tables. The results are given in Table VI.

We are now ready to carry out the summation of these terms with respect to k as indicated in (33). This means summing up the entries in each row of Table VI; the results of which are recorded in the sixth column, under the heading Σ_k.

TABLE VI

COMPUTATION OF $C_k^{26}\, C_{r-k}^{26}\, C_{4-k}^{26-r+k}\, C_{4+k-r}^{26-k}$

r	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	Σ_k
0	2.23502^8					2.23503^8
1	8.55140	8.55140^8				1.71028^9
2	1.12237^9	3.57604^9	1.12237^9			5.82078
3	5.98598^8	5.13084	5.13084	5.98598^8		1.14588^{10}
4	1.09359	2.99299	8.04609	2.99299^9	1.09359^8	1.42508
5		5.98598^8	5.13084	5.13084	5.98598	1.14589
6			1.12237	3.57604	1.12237^9	5.82078^9
7				8.55140^8	8.55140^8	1.71028
8					2.23503	2.23503^8

At this stage of the computation k disappears; that is, the numbers denoted by Σ_k depend upon r only. Therefore the remainder of the computation can be carried out in a single table.

The first column in this table (Table VII) contains the values of C_r^{26}. The next column contains the products $C_r^{26}\, C_{8-r}^{26}$, which are obtained by multiplying together the first and last entries in the first column, the second from the top and the second from the bottom, the third from the top and the third from the bottom, and so on. It is obvious that this column is symmetrical about the center of the column. Hence it is only necessary to write the first five numbers.

The third column contains C_r^8. It is now necessary to multiply Σ_k as given in Table VI by the corresponding number in the second column of Table VII, and divide it by the corresponding entry in the third column of Table VII. In using a modern computing machine it is easier to carry out both of these operations without writing down the intermediate result than otherwise. Therefore, the next column in Table VII contains the result of both operations, that is,

$$\frac{C_r^{26}\, C_{8-r}^{26}\, \Sigma_k}{C_r^8}.$$

As Σ_k and C_r^8 also are symmetrical about the middle of the column, all these computations are carried out for the top five entries only.

Finally (33) requires the addition of the numbers in this last column for every value of r from 0 to 8. The sum thus obtained is written at the bottom of the column. $P(n)$ is then obtained simply by dividing this sum Σ_r by the square of C_8^{52} as taken from Appendix III. This completes the solution of the problem.

TABLE VII

FINAL COMPUTATION OF THE DESIRED PROBABILITY

r	C_r^{26}	$C_r^{26}\, C_{8-r}^{26}$	C_r^8	$C_r^{26}\, C_{8-r}^{26}\, \Sigma_k / C_r^8$
0	1	1.5623^6	1	3.4917^{14}
1	26	1.7103^7	8	3.6563^{15}
2	325	7.4825^7	28	1.5555^{16}
3	2,600	1.7103^8	56	3.4996^{16}
4	14,950	2.2350^8	70	4.5501^{16}
5	65,780			
6	230,230			
7	657,800			
8	1,562,275			

$\Sigma_r = 1.5461^{17}$

$(C_8^{52})^2 = 5.6631^{17}$

Answer $P(4) = 0.27302.$

§ 31. *Some Instructive Illustrations; Another Urn Problem*

An urn problem of somewhat different type from those which have been previously considered is the following:

EXAMPLE 32.—*An urn contains m black and n white balls. These are drawn out one at a time and placed in a separate container. The drawing is continued until all those balls which remain are of the same color. What is the chance that they are all black?*

Suppose the problem were modified by requiring the drawing to continue until only one ball remained. If this were done, what would be obtained would be some one of the possible permutations of m black and n white things, the remaining ball being the last member of the permutation. The number

of such permutations in which the last ball is black is obviously equal to

$$P^{m-1,\,n}_{m-1,\,n},$$

and as the permutations are all equally likely to appear the chance that the last remaining ball is black is

$$\frac{P^{m-1,\,n}_{m-1,\,n}}{P^{m,\,n}_{m,\,n}} = \frac{m}{m+n}.$$

This same result could have been obtained by noting that any of the $m + n$ balls is equally likely to be left as any other one, and m of them are black.

This is also the answer to the problem as originally stated, a fact which becomes evident upon noting that whenever the last ball on this sort of drawing is black the group of balls which remain after the contents of the urn has been so far reduced as to be all of one color must of necessity be black. On the contrary, if the last ball happens to be white, the residual contents of the urn must likewise be white.

§ 32. *Some Instructive Illustrations; Another Urn Problem*

The solution of the following problem—like that of the last—involves no great difficulty if good judgment is used in the choice of the method of solution, but it might be very hard otherwise.

EXAMPLE 33.—*An urn contains m white balls and n red balls, m being greater than n. These are drawn out one at a time and placed in a second container. The drawing is continued until all the balls have been transferred. What is the chance that throughout this process there are always more white than red balls in the second container?*

It is obvious that any drawing will result in some one of the $P^{m,\,n}_{m,\,n}$ possible permutations of m white and n red balls. These permutations are all equally likely and form the complete group necessary for the application of the fundamental definition of probability.

To find the number of those permutations which satisfy

the condition that there are always more white than red balls in the second container is the next step in the solution. This can best be done by finding the number of permutations which do not satisfy this condition and subtracting that from the original number.

Any permutation which does not satisfy the conditions of the problem must belong to one of two classes: the class in which the first ball drawn is red, or the class in which the first ball drawn is white. These two classes can be considered separately.

Obviously, every permutation beginning with a red ball violates the condition of the problem; because, after the first ball is drawn, there are more red than white balls in the second container. The number of such permutations, however, is simply the number of ways in which the remaining $n - 1$ reds and m whites may appear; that is, $P_{m,n-1}^{m,n-1}$.

Some of the permutations which begin with white balls satisfy the conditions of the problem and some do not. At some stage of the process those which do not must have either the same number or a greater number of red than white balls in the second container. If there is a greater number, then at some subsequent stage the number must be equal, because when the drawing is complete the number of white balls exceeds the number of red. Therefore, finding the number of permutations which at some time have an equal number of red and white balls is the same as finding the permutations which violate the conditions of the problem.

This is most easily done by an artifice. The permutations under discussion are of the general type

w w r w r r | r w | w r | w w w.

At the stage in the drawing indicated by each of the lines there are the same number of red and white balls in the second container. This may occur more than once, as in the permutation given, but however many times the numbers may become equal there must always be a last time. Suppose, now, that for purposes of argument a new permutation is

built up which, after the last equality is reached, is identical with the one under consideration, but before the last equality is reached has red balls where the white ones were and white ones where the red ones were. In the case of the above illustrated permutation, for instance, this would lead to

r r w r w w | w r | r w | w w w.

This again is a possible permutation of m white and n red balls. Also it begins with a red ball. Hence, to every permutation which begins with a white ball and violates the conditions of the problem there corresponds by this inversion process a permutation which begins with a red ball. Furthermore, whenever the permutation beginning with a white ball is changed a different permutation beginning with a red ball is obtained. From this it follows at once that the number of permutations which begin with white balls and violate the conditions of the problem cannot exceed the number of permutations which begin with red balls. This can be expressed symbolically by saying that

$$p_w \leqq p_r.$$

Every permutation which begins with a red ball must somewhere have an equal number of reds and whites, since after the drawing is completed there are more whites than reds. Therefore, the process which was carried out above can be inverted, leading to the conclusion: For every permutation which begins with a red ball there is a permutation beginning with a white ball and violating the conditions of the problem. This means, of course, that the number of permutations beginning with red balls cannot exceed the number of permutations beginning with white balls and violating the conditions of the problem; or in symbols,

$$p_r \leqq p_w.$$

When this inequality is compared with the one above it is found that both can be satisfied only when $p_r = p_w$. But p_r has already been found to be $P^{m,n-1}_{m,n-1}$; hence it follows at once

that the total number of permutations which violate the conditions of the problem is

$$2\,P^{m,\,n-1}_{m,\,n-1}.$$

Those which remain form the subgroup which is required by the fundamental definition of probability. Their number is

$$P^{m,\,n}_{m,\,n} - 2\,P^{m,\,n-1}_{m,\,n-1}.$$

Therefore the answer to the problem is obtained in the form

$$\frac{P^{m,\,n}_{m,\,n} - 2\,P^{m,\,n-1}_{m,\,n-1}}{P^{m,\,n}_{m,\,n}} = \frac{m-n}{m+n}.$$

The last two examples have been given to illustrate the extent to which it is sometimes necessary to augment the routine processes of probability theory by common-sense methods in obtaining solutions of comparatively simple problems.

PROBLEMS

1. Make use of the results of Problems 5, 6 and 7 of § 21 to obtain the solution of Example 24.

2. In the first paragraph of § 23 it is stated that the unconditional probability of any black card being incorrectly called is $\frac{1}{2}$. Since there are four black cards the chance that either the first or the second or the third or the fourth is incorrectly called works out to be $\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}=2$. What is wrong with this argument?

3. If p +'s and n −'s are distributed at random along a line, what is the chance that no two −'s are adjacent?

4. By the use of (25) prove that $\sum_{p=0}^{r} C^m_p\, C^n_{r-p} = C^{m+n}_r$.

5. Show that Convention II, which led to (21), makes possible the solution of Example 16.

References for Outside Reading

1. Coolidge: *Probability*, pp. 13–25.
2. Chrystal: *Algebra*, pp. 571–586.
3. Todhunter: *Algebra*, pp. 447–461.
4. Bachelier: *Calcul des Probabilités*, pp. 1–8.
5. Czuber: *Wahrscheinlichkeitsrechnung*, Vol. I, pp. 28–72.

CHAPTER IV

Probability and Experiment; Bernoulli's Theorem

§ 33. *Introductory Remarks*

The reader who has dealt at all with statistics will already have remarked that nothing whatever has been said about the frequency with which an event will happen. It has been said that in tossing a penny heads is "equally likely" to appear as tails; never that "heads will appear as often as tails." The reason is, that the latter statement is not true. Try it and see. It is not even true "in the long run," unless that phrase is given a rather unusual shade of meaning which is dangerously near begging the question.

Neither is it true that "in a large number of independent *runs*, heads will as often exceed tails as tails will exceed heads"; for that is merely shifting attention from the event "a single throw" to the event "a run," and weak logic is never made stronger by obfuscation. Again the answer is, "Try it and see."

If these things were true, probability would be an experimental science, which I am quite convinced it is not. It is true that the outcome of an experiment may change the probability that a penny is bad—any accretion of pertinent information does that. If a large number of throws showed twice as many tails as heads, and we were carrying the weak end of a bet, we would probably insist on changing the penny. It would be more probable, because of the experiment, that the penny was loaded than it was before the experiment was performed. We may even have undertaken the experiment with a view to finding out whether the penny is good or bad; and if bad, how bad. But if so, in a strict logical sense, we

will never find out. Not only will we never find what the probability of tails is (that is, *how bad* the penny is); we can never even answer with finality the question, Is it bad?; for our result, no matter how one-sided it may be, would not have been impossible with a good penny. We will be led to a *presumption* that the penny is bad, and even to a *presumption* that tails is *approximately* so-and-so much more likely than heads; but to nothing more.

Why these things are true—that is, why we cannot determine the magnitude of a probability from experiment, and why we can nevertheless use experiment as a practical means for approximate evaluation of probabilities—is, broadly speaking, the subject of this chapter. Naturally the answer can best be given after suitable foundations have been laid for it. So with the question in our minds we will proceed to the foundation-building. For this purpose certain rather elementary, but perhaps not very familiar, mathematical ideas must be introduced.

§ 34. *Limits and Things which Approach Them*

Let us take, as a very simple example, the function $y = x^2$, or its graph (Fig. 4). We say that "y approaches the limit zero as x approaches zero." Just what do these words mean?

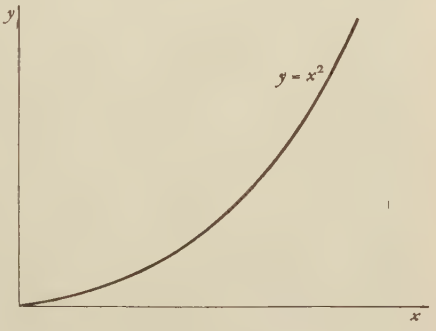


Fig. 4.

Do they mean that y is always zero? Certainly not; y is the ordinate to the curve. Do they mean y is ever zero? I suppose, since y *is* zero when x is zero, that the temptation may be to answer this question affirmatively: if it is not, so much the better. But we will suppose that temptation to exist, for the sake of overcoming it by another example—a trite one.

A man has a thread a foot long. He cuts off half of it. Then he cuts off half the remainder, and so on. What he has left after successive cuts then gives us the sequence

$$1, \quad \tfrac{1}{2}, \quad \tfrac{1}{4}, \quad \tfrac{1}{8}, \quad \tfrac{1}{16}, \quad \tfrac{1}{32}, \quad \cdots$$

Again, his remnant "approaches zero as time goes on." Is it ever zero? Obviously not.

There is, then, a real difference between the thing which approaches a limit, and the limit which it approaches. There may be special circumstances when the two merge, as in our first example, or there may not, as in our second.

It may even happen, that at just the point where we expect the variable to reach its limit it misbehaves and takes a different value. Consider, for example, Fig. 5, in which a function of x is represented which obviously approaches the limit $+1$ as x approaches zero from the left, and -1 as it approaches zero from the right. It is not quite apparent what to expect it to do *at* $x = 0$. But such a function can be represented by an infinite series of sine functions—a Fourier Series—and when it is so represented we learn the surprising fact that at $x = 0$ the function takes the value zero—a value which it certainly does not approach from either side.[1]

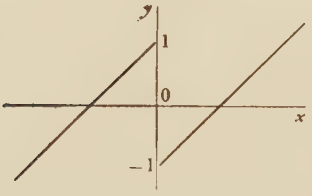


FIG. 5.

There is a further distinction between our examples: Figs. 4 and 5 define y for any value of x: for $x = 0.1$, for example. But in the thread illustration "the length of the remnant after one-tenth of a cut has been made" is nonsense. Calling N the number of cuts already made, and L the length of the remnant, this illustration would lead to a graph like that of Fig. 6, with ordinates at every integer point along the

[1] This result is a consequence of using the Fourier series, not an inherent property of two-line segments that fail to meet. The point is, not that the function *must* so behave, but that it *may*.

N-axis, and nothing in between. This is expressed by saying that the values of y in Figs. 4 and 5 constitute a " continuum," and those of Fig. 6 a " discrete set." Obviously either kind can approach a limit; and either kind can fail to reach that limit.

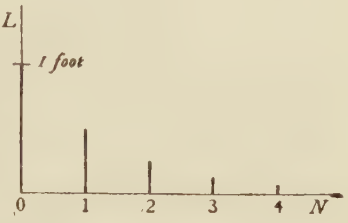


Fig. 6.

It happens that we are to be interested primarily in discrete sets: hence we phrase our definition of " approaching a limit " in the way most appropriate to that use. It is as follows:

If an ordered set of numbers is of such a nature that, having chosen in advance a number ϵ as small as we like, but not zero, and having cancelled some finite number of numbers from the set, we can assert that no two of the remainder differ by as much as ϵ, the set has a limit.

Let us take, for example, the set

$$1, \tfrac{1}{2}, \tfrac{1}{4}, \ldots$$

If a particular value of ϵ is chosen, there is some power of $\frac{1}{2}$ smaller than ϵ. Thus, if ϵ is 0.000001, $(\frac{1}{2})^{20} < \epsilon$. But whatever ϵ is, some power N can be found such that $(\frac{1}{2})^N < \epsilon$, this N being a finite number. If we cancel the first N terms of our set, no two of the remainder differ by as much as ϵ. Hence the set approaches a limit.

There is a defect to this definition: it does not tell us what " the limit " is. It merely tells what we mean by approaching a limit. This defect can be removed in one or the other of two ways:

(*a*) If the set contains a number q which need not be cancelled no matter how small ϵ is made, q is the limit.

(*b*) If to the set a new number q is added, and it is found that q need not be cancelled no matter how small ϵ is made, q is the limit.

Obviously, in the first case the set contains its limit or, in

technical language " is closed "; while in the second case the limit is not contained in the set, or the set " is open."

Our example is a case of an " open set "; for if we try *any* member of the set, say, $(\frac{1}{2})^m$, we can choose ϵ so small that it must be excluded: for example, $\epsilon = (\frac{1}{2})^{m+1}$ would do. But if we arbitrarily include zero in the set, zero need never be cancelled. Hence our set is an " open set " with the " limit zero " which it " does not contain."

§ 35. *The Upper Bound of a Set*

Among a *finite* set of numbers, one possesses the property of being at least as big as any other. It is the " largest." Sometimes there are several which possess this property. If so there are several " largest " numbers, as in the set 1, 2, 7, 7, 7.

This statement is only safe of finite sets, however. Among an infinity of numbers there need not be any largest number. There are two ways in which this rather unexpected result can come about. They can be very simply illustrated by the following examples:

There is no largest integer among the infinity of positive integers, for no matter what integer may be chosen there is always one larger than this. This is a case where the numbers contained in the set do not have any limit as regards size. That is, they are not " bounded."

The set of fractions $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots$, all of which are of the form $m/(m+1)$, does not contain a largest fraction; for if it did, this fraction could be obtained by giving m some particular value M. Whatever M may be, however, $m = M + 1$ always results in a larger fraction: that is, $(M+1)/(M+2)$ is bigger than $M/(M+1)$. Hence there is no largest fraction. This is a case where the set of numbers is actually " bounded "; there is no number bigger than 1. If, however, the integer 1 is included in the set, it is the biggest number. The set is now " closed " whereas before it was " open." In either case 1 is called the " least upper bound " of the set, the only difference being that the closed set contains its least upper bound (which is also the limit of the sequence); the open set does not.

Among numbers which represent probabilities there can be no unbounded sets, for such numbers are inherently contained between zero and one. But there may be either open or closed sets. In fact, the sequence $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots$, is capable of interpretation in terms of probability. If an infinity of urns are provided each containing one white ball, and if in one of these urns is placed one black ball, in another two black balls, in a third three black balls and so on, the chance that a ball drawn at random from one of the urns is black is either $\frac{1}{2}$ or $\frac{2}{3}$ or $\frac{3}{4}$ or $m/(m+1)$, according to the urn from which the ball is drawn. It is obvious from an intuitive standpoint that there is no urn in the set for which the probability is as great or greater than for all the remaining urns. The set is "open."

If, however, an additional urn is provided with black balls only, *this* will have the probability 1, which is actually greater than for any other urn. The set is now "closed."

In either of these cases 1 is called the "least upper bound" of the probability. In the second case it is reached (by the urn with only black balls), in the first case it is not.

It is a general theorem regarding sets of numbers that a closed set always contains its least upper bound (which is then a synonym for "largest number"), while an open set may, but need not. An example of an open set which does *not* has already been given. An open set which does—and which can easily be interpreted in the sense of probability—is $1, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{4}{9}, \ldots$, which does not contain its limiting value $\frac{1}{2}$, though it does contain a number 1 greater than any other number in it.

Now it happens that almost all statistical studies are postulated upon open sets which approach their upper bound as a limit but do not contain it, and a great deal of logical confusion has arisen from loose thinking about them. Hence it is essential to obtain a clear picture of the significance of the upper bound in such cases.

As a step in this direction, take the set of events already mentioned: the drawing of a black ball from the various urns

each having just one white ball. The set of numbers is $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$, . . . , and its upper bound (which is not reached) is 1. Can it be asserted that, by choosing wisely the urn from which we draw, we can be assured of obtaining a black ball? It cannot. So long as the urn contains a white ball—as all the urns do—the outcome is unknown. The most we can say is, that by passing from urn to urn the importance of our ignorance can be made less and less. The limiting condition is certainty, but that limit cannot be reached.

Choose your urn: make it as far along in the sequence as you like. I'll take the first one. We draw. We want black, but neither is certain to get it. *The only difference lies in the importance of our states of ignorance; yours is of less moment than mine.*

" But," you say, "if we draw repeatedly, my superiority will manifest itself." I have already said weak logic cannot be strengthened by obfuscation; but I'll follow you this once into what, if we kept on, would soon become an infinite regression. Name your number of trials. Make it as big as you like. You cannot be sure you will not draw white *every time*, and I black. But we'll agree it is extremely improbable.

What I am aiming to make clear is that " extremely improbable " is as far as we can go; and that *that* is not arrived at by trial, but by judgment in advance. A trial can only tell us what has happened; our intelligence can go further than that and say whether it was miracle or not.

§ 36. *Regarding Probability as a Limit*

There are certain students of the subject who define the probability of a head appearing when a penny is tossed, as the limit of the ratio of the number of heads to the number of throws, as the number of throws is increased indefinitely. Such a definition, however, implies as a fundamental postulate that the ratio obtained in this way *actually approaches a limit.* If this were so, such a definition would be logically possible, and I think would have considerable superiority over the one which we have set up; for we have already called attention

to the fact that there are many situations to which our definition cannot be applied, whereas the limit definition could be applied equally well[1] to almost any situation. The trouble is, that the fundamental postulate is only tenable provided the trials are *not independent*,[2] whereas the definition implies, even when it does not explicitly state, that the trials *are independent*.

To see how this inconsistency arises, let us consider the matter of tossing pennies, and let h_n denote the number of heads observed in n tosses. We form, in particular, the sequence of ratios,

$$\frac{h_1}{1}, \frac{h_2}{2}, \frac{h_3}{3}, \frac{h_4}{4}, \ldots, \frac{h_n}{n}, \ldots,$$

stretching out toward infinity as throw after throw is made. The definition which we are criticizing affirms that this sequence approaches a limit. But we can only properly make such an affirmation provided, after a finite number of terms have been eliminated, we are assured that no two of the remainder differ by more than some pre-assigned quantity ϵ. In fact, it must be possible to make this assertion for *any* ϵ, no matter how small; but for our purposes it is quite sufficient to consider only the value $\epsilon = \frac{1}{4}$.

Let the last of the terms which are to be eliminated be the one corresponding to $n = N - 1$, N being, of course, any number we please. Then, if a limit exists, it must be possible to assert that *no two* of the terms which remain differ by as much as ϵ. If there is any pair about which this assertion cannot be made, it is impossible to say that the sequence approaches a limit.

Consider, now, the particular terms corresponding to

---

[1] And also equally badly. In theoretical discussions it would always be applicable, and in practice never, for we can never cause our number of trials to "approach infinity" in a practical sense. This difficulty is a superficial one, however. The utility of experiment is exactly the same, no matter which point of view we start from.

[2] Unless "approaches a limit" means something logically different from its usual mathematical definition.

$n = N$ and $n = 2N$. If h_N/N exceeds $\frac{1}{2}$, and if it should happen to be true that every throw from $N + 1$ to $2N$ yielded a tail, it would be true that $h_{2N} = h_N$. Hence

$$\frac{h_N}{N} - \frac{h_{2N}}{2N} = \frac{1}{2}\frac{h_N}{N} > \frac{1}{4} > \epsilon.$$

Obviously it cannot be asserted that this will not happen, unless the result of a particular throw depends on what has gone before.

On the other hand, if h_N/N does not exceed $\frac{1}{2}$, and if the throws from $N + 1$ to $2N$ should all yield heads, it would be true that $h_{2N} = h_N + N$, whence

$$\frac{h_{2N}}{2N} - \frac{h_N}{N} = \frac{1}{2} - \frac{1}{2}\frac{h_N}{N} \geqq \frac{1}{4} \geqq \epsilon;$$

and it cannot be asserted that this will not happen, either. It is therefore impossible to say that the sequence has a limit.

The trouble lies, fundamentally, in the fact that the set of probabilities corresponding to a run of one head, two heads, three heads, and so on, though it has the limit zero, never reaches that limit. If it ever did reach it and remained there thereafter, the ratio of heads to throws would also have a limit. This fact would then be a theorem, not a postulate. But so long as the throws are independent, the next throw after a long run of heads may also be a head: no length of run is impossible, and the ratio *need not* (though it very probably will) *approach a limit.*

Or, if we prefer to look at it from a slightly different point of view, the trouble lies in the fact that the two positive assertions, "The trials are independent," and "The sequence approaches a limit," are inconsistent, and cannot be made the basis for a definition.

What can be done is this: an *a priori* definition of probability being allowed, it can be *proved* that the PROBABILITY of any two terms differing by more than ϵ can be made as small as desired

by taking N large enough. *It is the a priori probability, not the sequence of experimental ratios, which has a limit.* This will become more apparent in the course of the next few sections.

§ 37. *Regarding Repeated Independent Trials*

Consider an urn in which three balls are placed: one white and two black. Suppose repeated drawings are made from this urn, the balls being returned after each. According to (23), § 25, the chance of *just n white balls in m trials is*

$$P_m(n) = C_n^m \left(\tfrac{1}{3}\right)^n \left(\tfrac{2}{3}\right)^{m-n}.$$

Suppose now that five trials are made in succession. Either one of six things may happen: None of the trials may give a white ball, or one of them, or two, or three, or four, or five. The probability of each of these six results can be computed, the results being given in the accompanying Table VIII.

In this case there are two "most probable" results. One white ball or two white balls are equally likely to appear, and either is more probable than any other possible result.

TABLE VIII

THE PROBABILITY OF n SUCCESSES IN FIVE TRIALS IF THE PROBABILITY IN A SINGLE TRIAL IS $\frac{1}{3}$

n	Probability	n	Probability	n	Probability
0	0.1317	2	0.3292	4	0.0412
1	0.3292	3	0.1646	5	0.0041

If ten trials are made instead of five there are eleven possible results. The probabilities of these eleven results individually are given in Table IX. In this case there is one "most probable" number of white balls, the number being three. The least probable number of white balls is, of course, ten, for which the value given in the table is 0.0000. This does not mean that ten white balls could never appear in succes-

sion. It only means that the chance is so small as to be negligible in the fourth decimal place (that is, less than one-half of 0.0001). The exact probability is 1/59,049 = 0.0000169.

TABLE IX

THE PROBABILITY OF n SUCCESSES IN TEN TRIALS IF THE PROBABILITY OF SUCCESS IN A SINGLE TRIAL IS $\frac{1}{3}$

n	Probability	n	Probability	n	Probability
0	0.0173	4	0.2276	8	0.0030
1	0.0867	5	0.1367	9	0.0003
2	0.1951	6	0.0569	10	0.0000
3	0.2601	7	0.0163		

If fifty trials are made the results are as shown in Table X. Again there are two equally likely results, 16 and 17, each of which is more probable than any other.

TABLE X

THE PROBABILITY OF n SUCCESSES IN FIFTY TRIALS IF THE PROBABILITY OF SUCCESS IN A SINGLE TRIAL IS $\frac{1}{3}$

n	Probability	n	Probability	n	Probability
< 5	0.0000	13	0.0679	23	0.0202
		14	0.0898	24	0.0113
5	0.0001	15	0.1077	25	0.0059
6	0.0004	16	0.1178	26	0.0028
7	0.0012	17	0.1178	27	0.0012
8	0.0033	18	0.1080	28	0.0005
9	0.0077	19	0.0910	29	0.0002
10	0.0157	20	0.0704	30	0.0001
11	0.0287	21	0.0503		
12	0.0470	22	0.0332	>30	0.0000

The first and last entries mean, not that these cases cannot occur, but that their chances of occurrence are less than one in 20,000. As a matter of fact there is a finite probability for

drawing a white ball in every one of the fifty trials, but it is exceedingly small. In fact it is [1]

$$\frac{1}{3^{50}} = \frac{1}{717{,}898{,}000{,}000{,}000{,}000{,}000{,}000}.$$

§ 38. *The Limiting Condition as the Number of Trials is Greatly Increased*

The probabilities in Tables VIII to X are plotted as ordinates in Fig. 7.

In this form of graphical presentation several facts stand out at once:

1. There is an orderly progression from one graph to the next.

2. The most probable number of successes increases progressively; that is, the greater the number of trials the greater the most probable number of successes.

3. The probability of this most probable number of successes decreases progressively; that is, the greater the number of trials the less the chance of coming out with the most probable result.

4. The number of different results which have probabilities comparable to that of the most probable result increases progressively. This result can be expressed quite simply by saying: the greater the number of trials, the greater the spread of the chart. In other words: the probability of missing the most probable result by more than a stated amount increases constantly as the number of trials is increased.

This last point is worthy of some further discussion.

Suppose, for example, that we ask for the chance of missing the most probable result by more than five units. If only five trials are made it is impossible to miss the most probable result by more than five units. In this case the probability

[1] Numbers such as this are inconceivable. I suggest, as an entertaining exercise of the imagination which is worth while just once, the computation of the dimensions of a container which would hold 3^{50} beads. Having done this, if *just one* were black and the rest all white, and if the bunch were thoroughly mixed, the chance of drawing the black one would be the fraction in question. "Too small to matter," you say: yet the chance of getting the one you *did* get was no greater!

asked for is zero. On the other hand, if ten trials are made, the probability of missing the most probable result by more than five units, though quite small, is a finite value. It is[1] 0.0003. In the case of fifty trials the chance of missing the

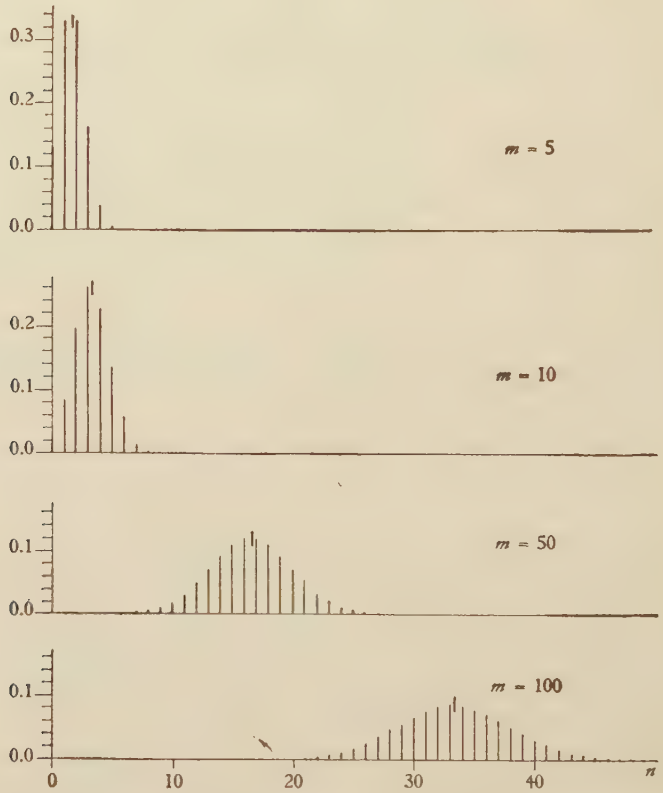


FIG. 7.—THE PROBABILITY OF n SUCCESSES IN m TRIALS IF THE PROBABILITY OF SUCCESS IN A SINGLE TRIAL IS $\frac{1}{3}$.

most probable result by more than five is 0.13. Moreover, this probability can be made as large as we please by taking

[1] This figure is obtained as follows: In Table IX the most probable value of n is 3. To miss it by more than 5 would require that n be less than zero, which is impossible, or more than 8, for which the probability 0.0003 is obtained by adding together the last two entries in the table.

The figures for 50, 100 and 1000 trials are obtained in a similar way.

the number of trials large enough. Thus, for 100 trials it is 0.29, for 1000 trials 0.74, and for 1,000,000 trials about 0.99.

Though these figures are based upon missing the most probable result by *five* units, the same qualitative facts apply to missing it by any preassigned number of units.

In an infinity of trials the probability approaches certainty, that the observed result will differ from the most probable result by more than any preassigned number, no matter how large.

It is this spreading out of results which is responsible for the tendency toward decreasing probability of the most probable result. Since one or the other of the various results is bound to happen, the sum of the ordinates on each of the graphs must of necessity equal unity. As the number of ordinates of comparable length increases, the magnitude of each must be necessarily decreased.

Put in still another form this statement becomes: The probability of missing the most probable result by more than five is the sum of all the ordinates which lie outside a band width of five on each side of the highest ordinate. Since this sum increases progressively as the number of trials is increased, it follows that the sum of all the ordinates within the band must correspondingly decrease.

5. The most probable number of successes is always approximately one-third the number of trials.

This is illustrated in the figure by a short vertical line drawn near the top of the curve for that value of n which corresponds to $m/3$. In the case of five trials $m/3$ is $\frac{5}{3}$ and lies between 1 and 2. These are the two most probable results. In the case of ten trials, $m/3$ is $\frac{10}{3}$ and lies between 3 and 4. The most probable number of successes is 3. For fifty trials the line comes at $\frac{50}{3}$ and it was found that 16 and 17 were each equally likely. For 100 the line comes at $\frac{100}{3}$ and 33 is the most probable number of successes.

Since the probability of success in a single trial is $\frac{1}{3}$, this suggests the rule that the most probable number of successes

in m trials is mp, where p is the probability of success in a single trial, provided mp is an integer; otherwise the most probable number of successes is one or the other of the integers between which mp lies. Although it is not always safe to generalize particular cases in this fashion, it happens in the present instance that the result is correct.

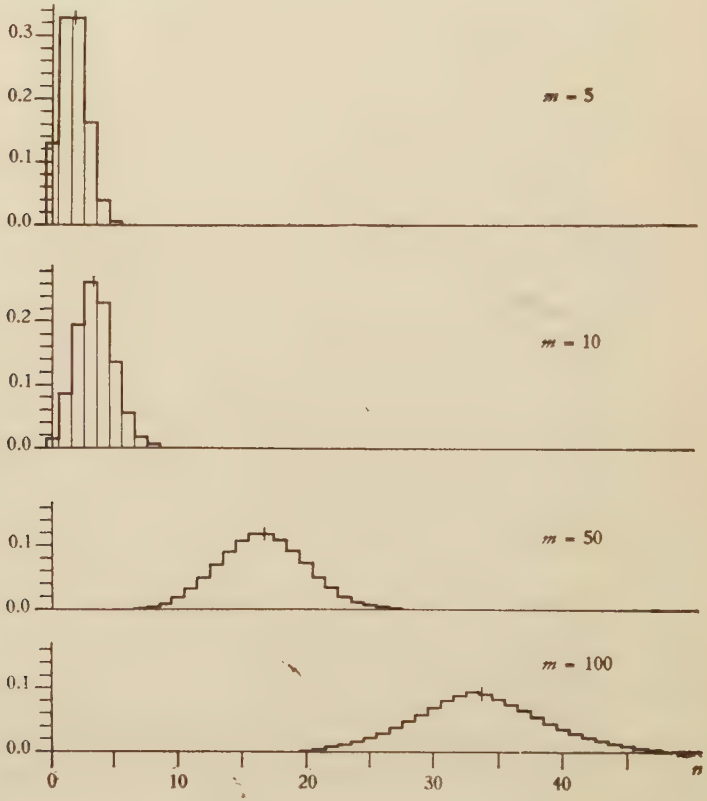


FIG. 8.—AN ALTERNATIVE FORM OF FIG. 7.

There is one more thing which it is worth while to do with Fig. 7. Since the ordinates which represent the probabilities occur at unit intervals while intermediate values of n have no significance, it is possible to erect a rectangle of unit width upon each ordinate, thus producing a set of rectangles the

areas of which are equal to the probabilities of the values of n upon which they stand.

When so treated the diagram takes the form shown in Fig. 8, in which the graphs corresponding to five trials and ten trials are drawn exactly as explained. The other graphs differ from these only in the fact that the vertical sides of the rectangles, which contribute nothing to the interpretation of the graphs, are omitted, leaving only a broken line. This broken line has two unique properties. One is the property from which it was derived—that the area under each step is the probability of the corresponding value of n. The other property comes from the fact that the set of values of n is complete; it is that the area under the entire broken curve is unity.

A curve constructed in this fashion is called a " distribution curve " for the variable n.

§ 39. *Bernoulli's Theorem*

The fact that the most probable number of successes in Figs. 7 and 8 is just about $\frac{1}{3}$ the number of trials suggests plotting the set of curves once more, using, however, n/m instead of n as abscissa. In doing this we shall obtain a sort of distribution curve for the proportion of successes, or as we shall more often call it, a " percentage distribution curve." But in order that the word " distribution curve " may have the same meaning as before, the two unique properties to which reference was made in § 38 will be conserved. That is, the *areas* of the rectangles will be kept constant, no matter what happens to the *ordinates*. Naturally, since the curve for $m = 100$, say, will be condensed more laterally than will the curve for $m = 10$, it will also be stretched more vertically, with the result that the curves will be differently related than before.

The reconstructed family is shown in Fig. 9 for $m = 50$, 100 and 1000.[1] Again there is a marked progressive tendency

[1] For $m = 1000$ the steps are so small that they cannot be shown.

among the curves, but the laws of progression are no longer the same as before. They may now be stated as:

1. The most probable proportion of successes remains approximately the same as m increases.

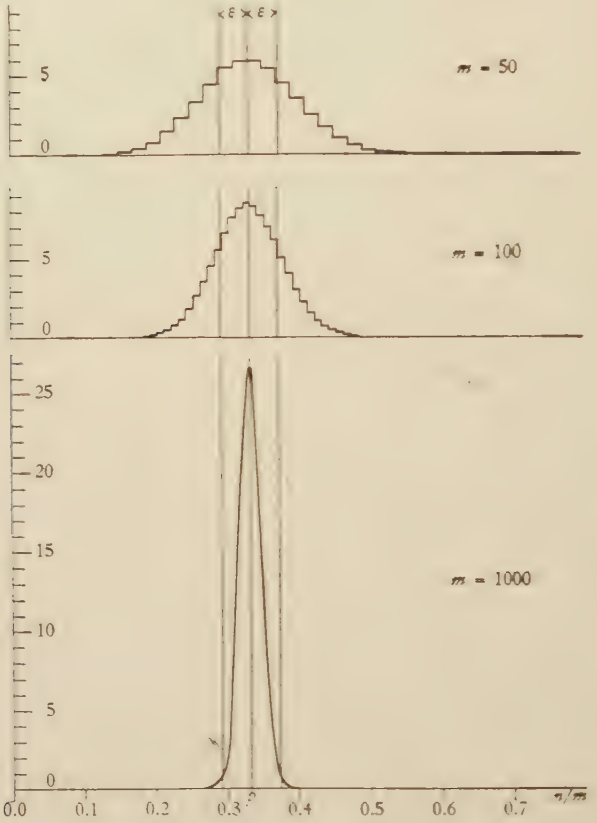


FIG. 9.—AN ALTERNATIVE FORM OF FIG. 7.

2. This most probable proportion is always as near to p as it can be, considering the fact that n must be an integer.

3. The height of the rectangle which represents this most probable proportion *increases* as the number of trials increases.

4. The spread of the percentage distribution curve *decreases* as the number of trials is increased. That is, although the chance

of missing the most probable value of n by more than a preassigned *amount* gets greater and greater as the number of trials increases—a fact to which attention was called in the last section—the chance of missing this most probable value of n by a given *percentage* decreases continually.

It is easy to see that the chance of n/m differing from $\frac{1}{3}$ by less than a preassigned amount ϵ is represented in Fig. 9 by the area bounded laterally by a pair of vertical lines[1] ϵ units to each side of $\frac{1}{3}$. In the figure ϵ is taken as 0.04. With only 50 trials more than half the area lies outside these limits. That is, the proportion of successes is more likely to lie outside the limits 0.293 and 0.373 than inside them. In the case of 100 trials the area outside the boundary is considerably reduced, and the proportion of successes is more likely to lie between the prescribed limits than not. In the case of 1000 trials almost the entire area lies between these limits: there is very little probability that in so extensive an experiment the proportion of successes would differ from $\frac{1}{3}$ by as much as 0.04. By increasing m still further, a point would eventually be reached where the chance of the proportion of successes lying outside the prescribed range would be smaller than any arbitrarily fixed quantity. In other words, as m approaches infinity, the chance of n/m lying outside the limits $p - \epsilon$, $p + \epsilon$ approaches zero, and the chance that it lies within these limits approaches certainty.

This tendency has been illustrated by the use of $\epsilon = 0.04$ and $p = \frac{1}{3}$, but the conclusion reached would be the same no matter what values were chosen. If, for example, ϵ were taken as 0.000001, the chance of n/m lying within $p - \epsilon, p + \epsilon$ would be extremely small for a hundred trials or even a thousand trials; but it would get larger and larger as the number of trials was increased until with some very large number of trials, say a million million, it would be almost certainty. This fact can be expressed in the form of a theorem as follows:

[1] Of course those "steps" which lie partly within this band and partly outside it are either to be entirely included, or else entirely excluded, according as their midpoints lie within or without the band.

BERNOULLI'S THEOREM: *If the chance of an event occurring upon a single trial is p, and if a number of independent trials are made, the probability that the ratio of the number of successes to the number of trials differs from p by less than any preassigned quantity, however small, can be made as near certainty as may be desired by taking the number of trials sufficiently large.*

Sometimes the content of a theorem such as this is made clearer by throwing mathematical discretion to the winds and stating it in the form of every-day language. The present appears to be a case of this sort, and therefore we restate the theorem as follows:

If the probability of an event is p, and if an infinity of trials are made, the proportion of successes is sure to be p.

This, of course, is exactly the statement to which we objected in § 34; yet the statement is as certainly "true" in one sense of the word, as it is *not* "true" in another. That it fails to stand the test of mathematical rigor, I believe the argument of § 36 shows. It is therefore not a fit foundation for a mathematical theory. But our every-day life is not conducted on such rigorous requirements as to "truth." You say, "Are you sure that he is coming tomorrow?" and receive the answer, "Yes." Both you and your informant understand what you mean: the event is contingent upon his not dying, for example, and perhaps on many other unforeseen circumstances. It is, in fact, not sure at all; it is merely very probable: so probable that the residual doubt is not worth expression. Our statement is in the same class. In fact, the residual doubts are even vastly smaller, and may quite properly remain unexpressed.

By painstaking experiment with the bad penny of § 33 we can learn the extent of the bias in favor of tails, in the sense that the chance of serious error is negligible. Or we can accumulate vital statistics and learn the chance of a man, about whose state of health we have no special information, dying at forty, with quite enough assurance for the purposes of a life

insurance company. But we should not be unaware of the logical status of what we are doing.

I agree with the statistician who says, "Life insurance is no gamble. Its laws are as immutable as those which cause the sun to rise." In a sense, what he says is true. But I cannot agree with him when he tells me that the entire logical foundation of the Theory of Probability is to be found in the taking of statistical data.

§ 40. *Résumé*

It has seemed impossible to give a discussion of Bernoulli's Theorem without traveling rather far afield at times, and as these excursions have removed the emphasis somewhat from the facts upon which it should rest, it is probably desirable to put those facts together in a compact form. As they divide themselves into two sets, one concerned with the *number* of times an event occurs, the other with the *proportion* which that number bears to the number of trials, they will be listed in parallel columns.

Facts about the Number of Successes

(1) In m independent trials under the same essential conditions, the number of times an event occurs, n, may take any value from 0 to m.

(2) There is a "most probable" number of successes. (There may be two.)

(3) This most probable number is pm, when pm is an integer; otherwise it is one (or both) of the adjacent integers.

(4) The chance of the number of successes differing from the most probable number by less than a fixed amount, *no matter how large*, approaches *zero* as the number of trials is indefinitely increased.

Loosely: In an infinity of trials the difference between the actual number of successes and the most probable number will be infinite.

Facts about the Proportion of Successes

(1) In m independent trials under the same essential conditions, the proportion of times an event occurs, n/m, may take any value from 0 to 1.

(2) There is a "most probable" proportion of successes. (There may be two.)

(3) This most probable proportion is either p, or the integral multiple of $1/m$ next smaller than p, or the one next larger than p, or both.

(4) The chance of the proportion of successes differing from the most probable proportion by less than a fixed amount, *no matter how small*, approaches *unity* as the number of trials is indefinitely increased.

Loosely: In an infinity of trials the difference between the actual proportion of successes and the most probable proportion will be zero.

§ 41. *Mathematical Justification*

So far in this discussion we have avoided formal mathematics entirely, basing our argument on common-sense inferences rather than proofs. We must now correct that defect. To start with we shall prove that the most probable number of successes is either mp or an adjacent integer, by the process of showing that each ordinate of Fig. 7 is bigger than the preceding one up to mp, and less than the preceding one from there on.

If we take two adjacent values of n, say, $n-1$ and n, and divide the probability of the second by that of the first, we obtain

$$\frac{P_m(n)}{P_m(n-1)} = \frac{m-n+1}{n}\,\frac{p}{1-p} = \frac{1-\frac{n}{m}+\frac{1}{m}}{\frac{n}{m}}\,\frac{p}{1-p}.$$

Now, reducing the numerator of the last member (by cancelling $1/m$) makes this member too small. Hence

$$\frac{P_m(n)}{P_m(n-1)} > \frac{1-\frac{n}{m}}{1-p}\cdot\frac{p}{\frac{n}{m}}. \tag{34}$$

Similarly, reducing the denominator by inserting -1 m makes the last member too large. Hence

$$\frac{P_m(n)}{P_m(n-1)} < \frac{1-\frac{n-1}{m}}{1-p}\,\frac{p}{\frac{n-1}{m}}. \tag{35}$$

To establish our theorem we need only notice that, if $n \leqq pm$, both factors on the right of (34) are 1 or greater, and therefore $P_m(n) > P_m(n-1)$. But if $n-1 \geqq pm$, (35) gives $P_m(n) < P_m(n-1)$.

These results, interpreted graphically, say that each

ordinate of Fig. 7 is bigger than the preceding one *up to the stroke at mp* (and including the ordinate at *mp* if there is one); and similarly that *beyond mp* every ordinate is smaller than the preceding one. Hence the first ordinate past the stroke is bigger than all that follow, and the last one before the stroke is bigger than all that precede. Either one of these is larger than the other, in which case it is a maximum; or else they are equal, in which case they constitute a pair of maxima.

Thus we have justified points (2) and (3) of our résumé. That the argument applies to Fig. 9 as well as to Fig. 7 is obvious, since they are identical except for scale. Hence the proof includes both columns of the résumé.

§ 42. *Stirling's Formula*

To justify the statements (4) we need some means of estimating the value of the expression $C_n^m p^n (1-p)^{m-n}$ when m is very large and n is nearly equal to mp, which usually means that it is very large too. Under these circumstances, however, C_n^m contains three factorials of very large numbers, and we have already seen in § 10 that such quantities become almost inconceivably large. For this investigation we need a new tool: an approximate formula for $n!$ when n is large.

In this case, in speaking of an "approximate formula" we do not imply that the difference between the true value and the approximation is small; it is the percentage error which is of consequence. For example, a formula which gave a result 3.0415×10^{64} for 50! would be a good approximation, for it differs from 50! only in the fifth place. But the actual magnitude of the difference is nevertheless a very large number. It is of the order of magnitude of 10^{60}.

In § 11 we adopted the definition

$$m! = \int_0^\infty x^m e^{-x}\, dx, \tag{6}$$

which, interpreted graphically, means that factorial m is the area included between the curve

$$y = x^m e^{-x}$$

and the x-axis. We get our first indication as to how to proceed in attempting to find an approximation formula by considering a few of the curves that arise from various values of m. It appears

from Fig. 10*a* that when m is large the principal contribution to the value of the factorial is made by that part of the curve in the neighborhood of the maximum ordinate. The " tails " in the neighborhood of $x = 0$ and $x = \infty$ dwindle away very rapidly and probably contribute little to the actual integral. This suggests that the area

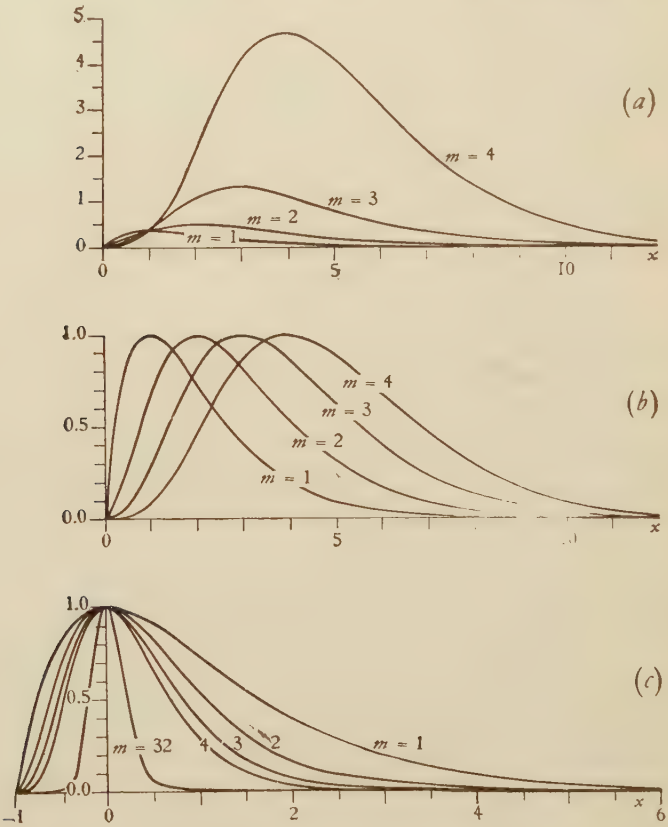


FIG. 10.—ILLUSTRATING THE CHANGES OF VARIABLE BY MEANS OF WHICH STIRLING'S FORMULA IS OBTAINED.

itself will be roughly proportional to the maximum ordinate, at least in the sense that the ratio of the two will not vary with m in such an extreme fashion as does the factorial itself.

It is a simple matter to determine this maximum ordinate Y. It turns out to occur at the point $x = m$, and to have the value

$$Y = m^m e^{-m}.$$

Dividing (6) by Y, and denoting the ratio by $f(m)$, we have

$$f(m) = \frac{m!}{Y} = \int_0^\infty (x/m)^m \, e^{m-x} \, dx.$$

This integral can also be interpreted as the area under a curve: the particular curve, of course, being identical with the corresponding curve in Fig. 10*a*, except that the vertical scale has been reduced in such a way as to make the maximum ordinate unity. Such a set of curves is shown in Fig. 10*b*.

We now shift these new curves so that the maximum ordinate of each lies on the y-axis (which means that we replace x by a new variable $x' = x - m$), and then contract them in the direction of the x-axis in the ratio 1 to m (by replacing x' by a new variable $x'' = x'/m$). The result of the two substitutions is

$$f(m) = m \int_{-1}^\infty [(x'' + 1) \, e^{-x''}]^m \, dx'';$$

or, since the primes are of no further service,

$$f(m) = m \int_{-1}^\infty [(x + 1) \, e^{-x}]^m \, dx. \tag{36}$$

The curves corresponding to this integral are shown in Fig. 10*c*.

The next step in our process is an ingenious one, only to be justified by the fact that it works. We replace x by a new variable u, defined by the equation

$$(x + 1) \, e^{-x} = e^{-u^2}; \tag{37}$$

or what amounts to the same thing,

$$x - \log (x + 1) = u^2.$$

Then we have

$$dx = 2u \, du + 2 \frac{u}{x} \, du,$$

and (36) breaks up into the sum of the two terms

$$f(m) = m \int e^{-mu^2} \, 2u \, du + 2m \int e^{-mu^2} \frac{u}{x} \, du.$$

For the moment we disregard the question of the limits of integration.

The first term in this expression integrates immediately into $-e^{-mu^2}$; that is, into $-[(x + 1) \, e^{-x}]^m$. As this quantity vanishes for both $x = -1$ and $x = \infty$, the first term is zero. This leaves us finally with the integral

$$f(m) = 2m \int e^{-mu^2} \frac{u}{x} \, du. \tag{38}$$

There is now nothing to do but to evaluate this integral directly. But to do this it is necessary to obtain an expression for x in terms of u, and as (37) is transcendental the solution can only be obtained as a series. It is found that [1]

$$\frac{u}{x} = \sqrt{2}\left(\tfrac{1}{2} + \tfrac{1}{12}u^2 + \tfrac{1}{432}u^4 - \tfrac{139}{194400}u^6 + \ldots\right) - \left(\tfrac{1}{3}u + \tfrac{4}{135}u^3 - \tfrac{4}{2835}u^5 + \ldots\right). \quad (39)$$

We must now determine the limits of integration. From (37) it follows that $u^2 = \infty$, both when $x = -1$ and when $x = \infty$. Hence u is $\pm\infty$ at both limits of integration. If the limits were equal the integral would vanish,[2] and we are sure from geometrical considerations that this cannot occur. It therefore follows that we must assign $-\infty$ as one limit and $+\infty$ as the other. If we were to assign them wrongly we would merely change the sign of the integral; and as we know that the value of the factorial must be positive we would detect that error at once. By trial it is found that $+\infty$ should be chosen as the upper limit and $-\infty$ as a lower one. Then, substituting (39) in (38) and integrating term by term, we arrive at the final result

$$f(m) = \sqrt{2\pi m}\left(1 + \frac{1}{12m} + \frac{1}{288m^2} - \frac{139}{51840m^3} + \ldots\right)$$

and therefore

$$m! = \sqrt{2\pi m}\, m^m e^{-m}\left(1 + \frac{1}{12m} + \frac{1}{288m^2} - \frac{139}{51840m^3} + \ldots\right). \quad (40)$$

This is known as Stirling's formula.[3] As an illustration of its use we may compute 100! which turns out to be

$$100! = \sqrt{200\pi}\; 100^{100}\, e^{-100}\left(1 + \tfrac{1}{1200} + \tfrac{1}{2880000} - \ldots\right)$$
$$= 9.332621\,5^{157}.$$

This is identical with the result given in Appendix I.

---

[1] The region of convergence of this series is $|u| < \sqrt{2\pi}$. We do not here attempt to justify its use with infinite limits.

[2] If u is defined by the principal branch of the logarithm, u^2 is positive for every value of x on the path of x-integration. Therefore, the u does not leave the real axis. As there are no singularities of u/x on the real axis, other than at $u = \pm\infty$, a path from infinity to a finite point and back could only return via the same branch of the function.

[3] The name is usually applied to the first term only:

$$m! = \sqrt{2\pi m}\, m^m\, e^{-m}.$$

The following table contains several additional factorials, together with their Stirling Approximations to one and four terms:

TABLE XI

COMPARISON OF THE FACTORIAL WITH ITS STIRLING APPROXIMATIONS

m	True Value	One Term		Four Terms	
	$m!$	$m!$	% Error	$m!$	% Error
1	1	0.922137	8	0.999711	0.03
2	2	1.919005	4	1.999986	0.0007
5	120	118.0192	2	120.0000	
10	3,628,800	3,598,696	0.8	3,628,800	
100	9.3326215^{157}	9.324847^{157}	0.08	9.3326215^{157}	

It will be noticed that for values of m greater than 10 even the simpler formula gives quite acceptable accuracy. For smaller values it would not be used because of the ease with which the factorials themselves may be found.

§ 43. *Another Approximate Formula*

We still need an approximation to $(1 + x/m)^m$ when m is very large. Obviously, when m is very large $1 + x/m$ is only slightly different from 1. Hence any moderate power would also be but slightly different from 1; but the mth power is not "moderate," so the difference may be considerable.

First let us find the limit approached as m becomes infinite. This we do by writing

$$\log\left(1 + \frac{x}{m}\right)^m = m \log\left(1 + \frac{x}{m}\right).$$

Obviously, since

$$\log\left(1 + \frac{x}{m}\right) = \frac{x}{m} - \frac{x^2}{2m^2} + \frac{x^3}{3m^3} - \cdots,$$

$$\log\left(1 + \frac{x}{m}\right)^m = x - \frac{x^2}{2m} + \frac{x^3}{3m^2} - \cdots.$$

Hence

$$\left(1 + \frac{x}{m}\right)^m = e^{x - (x^2/2m) + (x^3/3m^2) - \cdots},$$

which obviously approaches the limit e^x as m approaches infinity. Unity would therefore have been a very poor approximation. The fact that

$$\lim_{m=\infty} \left(1 + \frac{x}{m}\right)^m = e^x \tag{41}$$

is the first important result of this section.

If we like, we may think of e^x as a *first approximation* to the value of $(1 + x/m)^m$, when m is large. We can, however, obtain a better approximation by segregating the factor $e^{-(x^2/2m)+(x^3/3m^2)-(x^4/4m^3)+\cdots}$, and expanding it in a series of decreasing powers of m. As a result we find

$$\left(1 + \frac{x}{m}\right)^m = e^x\left[1 - \frac{x^2}{2}\frac{1}{m} + \left(\frac{x^4}{4} + \frac{2x^3}{3}\right)\frac{1}{2!\,m^2} - \left(\frac{x^6}{8} + x^5 + \frac{3x^4}{2}\right)\frac{1}{3!\,m^3} + \left(\frac{x^8}{16} + x^7 + \frac{13x^6}{3} + \frac{24x^5}{5}\right)\frac{1}{4!\,m^4} + \cdots\right]. \tag{42}$$

This is the second important result of the section.

§ 44. *Justification of the First Half of Bernoulli's Theorem*

We are now ready to justify Bernoulli's Theorem by proving two things:

(*a*) *That the maximum ordinate in Fig. 7 approaches zero as m becomes infinite.* As there are only a finite number of ordinates between $mp - \epsilon$ and $mp + \epsilon$, and as none exceeds the maximum, the sum of all of them must therefore approach zero, which proves half the theorem.

(*b*) *That in Fig. 9 the ordinates (except those in the immediate vicinity of $n/m = p$) decrease so rapidly as m becomes infinite that the sum of all those which lie outside the band ultimately vanishes.* This will prove the other half. The proof of (*a*) will be given in this section; the proof of (*b*) in the next.

We begin with the formula

$$P_m(n) = C_n^m\, p^n\, (1 - p)^{m-n}, \tag{23}$$

and replace all factorials by their Stirling approximations. The result is

$$P_m(n) = \sqrt{\frac{m}{2\pi n(m-n)}}\left(\frac{m(1-p)}{m-n}\right)^m\left(\frac{(m-n)p}{n(1-p)}\right)^n. \quad (43)$$

Now the most probable n does not differ from mp by more than a unit. Hence

$$mp - 1 < n < mp + 1,$$
$$m(1-p) - 1 < m - n < m(1-p) + 1.$$

As replacing the denominators by *smaller* numbers, and the numerators by *larger* ones, increases the right-hand side of (43) it is obvious that

$$P_m(n) < \sqrt{\frac{m}{2\pi(mp-1)[m(1-p)-1]}} \times\left(\frac{m(1-p)}{m(1-p)-1}\right)^m\left(\frac{m(1-p)+1}{mp-1}\frac{p}{1-p}\right)^n.$$

From this point on we consider the three terms separately.

First it is obvious that the first factor approaches zero with increasing m; for m occurs twice in the denominator and only once in the numerator. Hence if the other factors do not become infinite our proposition is proved.

The second factor, however, is of the form

$$\frac{1}{\left(1-\frac{1}{m(1-p)}\right)^m},$$

which reduces to (42) if $-1/(1-p)$ is called x. Hence this factor approaches $e^{-1/(1-p)}$, and does not become infinite.

The third factor can be rearranged so as to take the form

$$\left(\frac{1+\frac{1}{m(1-p)}}{1-\frac{1}{mp}}\right)^n.$$

As the quantity within the parentheses is obviously greater than unity, the factor will be increased if n is replaced by something larger than itself, as, for example, by $mp + 1$. Thus the third factor is less than

$$\frac{\left(1 + \frac{1}{m(1-p)}\right)^{mp}}{\left(1 - \frac{1}{mp}\right)^{mp}} \frac{\left(1 + \frac{1}{m(1-p)}\right)}{\left(1 - \frac{1}{mp}\right)}.$$

As m approaches infinity the last part of this product reduces to unity; while by identifying mp with the m of (42), and $p/(1-p)$ with x, the numerator of the first part becomes $e^{p/(1-p)}$; and by a similar process the denominator becomes e^{-1}. As none of these is infinite (43) must vanish. This proves our theorem.

§ 45. *Justification of the Second Half of Bernoulli's Theorem*

The second half of Bernoulli's Theorem is easier to justify. We know that the ordinate at a particular n is bigger than at any other n further removed from the most probable. Let us, then, consider an n equal to $m\eta$, where η is not exactly p. We have, from (43),

$$P_m(n) = \sqrt{\frac{1}{2\pi\eta(1-\eta)m}}\left[\left(\frac{1-p}{1-\eta}\right)^{1-\eta}\left(\frac{p}{\eta}\right)^{\eta}\right]^m,$$

or

$$P_m(n) = \sqrt{\frac{1}{2\pi\eta(1-\eta)m}}\, z^m,$$

if the bracketed expression is denoted by z.

Our first step is to show that z is always less than unity. This is best done by finding the maximum value of z: or rather of $\log z$, for that is easier. We have

$$\log z = (1-\eta)[\log(1-p) - \log(1-\eta)] + \eta(\log p - \log \eta).$$

By actual differentiation it is found that

$$\frac{d\log z}{d\eta} = \log\left(\frac{1-\eta}{1-p}\frac{p}{\eta}\right). \tag{44}$$

If this is to be zero—as it must be for a maximum value—η must equal p; and when $\eta = p$, $\log z = 0$. Hence the maximum value of z is unity. For any other value of η, $z < 1$.

Finally suppose η is set equal to that one of the quantities $p - \epsilon$, $p + \epsilon$ which happens to give the larger ordinate in Fig. 9. Then outside the range bounded by these quantities *all* the ordinates are smaller than the one at η: that is, they are all less than $\sqrt{1/2\pi\eta(1-\eta)m}\, z^m$. As there are less than m of them, their sum cannot exceed $\sqrt{1/2\pi\eta(1-\eta)}\,\sqrt{m}\, z^m$. But since $z < 1$, the quantity $\sqrt{m}\, z^m$ can be made as small as we like by taking m sufficiently large. In other words, the sum of all ordinates outside the range $p + \epsilon > n/m > p - \epsilon$ can be made as small as we wish by taking m large enough.

This proves the second half of our theorem. Later on, in § 82, we shall find a means of telling how great the probability is of either n or n/m lying in any such range, without the labor of computing the individual ordinates. We shall find, in fact, that if we were to draw another set of distribution curves for the variable $n/\sqrt{m}$, the separate curves of this set would be almost indistinguishable for large values of m. This is, in a way, suggested by the fact that that factor of (43) which vanishes as m increases indefinitely, does so because of the occurrence of $\sqrt{m}$ in the denominator. This suggests that the maximum ordinates of the curves of Fig. 8 would all be about equal if they were multiplied by the square roots of their respective values of m. But if the areas of the individual rectangles are not to be altered, such a change can only be brought about by condensing the curves laterally in the same ratio, which is equivalent to plotting the distribution functions for $n/\sqrt{m}$.

Though the present is not a suitable place to prove it, this is actually the case: As the number of trials is increased the peaks of Fig. 8 become lower and lower in proportion to the reciprocals of the square roots of the number of trials, while the curves spread out laterally to greater and greater extents, in proportion to the square roots of m.

§ 46. *Regarding the Experimental Determination of Probability*

Bernoulli's Theorem says, in substance, that the chance of an important difference existing between p, the probability of success in a single trial, and n/m, the proportion of successes in m independent trials, may be made as small as we please by making the number of trials large enough. If this is true, it is quite obvious that n/m may be accepted for most practical purposes as an approximation to the probability p; and this affords us a new way of measuring probability.

In the case of a perfectly good penny, for instance, we could either form the set of equally likely, mutually exclusive events "heads, tails," and conclude at once that the chance of obtaining a head is $\frac{1}{2}$; or we could toss the penny repeatedly and accept the proportion of successes as the value of the probability. In this specific case, of course, the former method would be by far the better, for it gives us an *exact* answer whereas the latter method gives only an approximation at best. But there are many questions to which this exact method cannot be applied, because no suitable set of alternative events is known. Take, for example, the chance of a man of twenty dying between the ages of fifty and fifty-five: it is quite out of the question to set up a complete set of equally likely events in this case. But it *is* possible to pick out a large number of men of age twenty, and by waiting twenty-five years determine what proportion actually die between the specified ages. If we are willing to admit that the chance of one of them dying is not influenced by what occurs to any of the others, we may accept this proportion as a satisfactory experimental determination of the desired probability.

Bernoulli's Theorem, therefore, furnishes us an acceptable makeshift when the direct a priori determination of probability is not feasible.

But we must observe caution in accepting this argument. Bernoulli's Theorem has been proved by the use of (23), and (23) was derived by the use of (20), which itself has only been established for cases to which the fundamental definition can

be applied directly. Hence either Bernoulli's Theorem itself, or the Multiplication Theorem (20), must be put upon a more general basis than has so far been done, if we are to use it as a justification of this experimental method of measuring probability. As the generalization of the Multiplication Theorem is not difficult so long as the events A and B are independent [as they are supposed to be in deriving (23)], we shall give this generalized proof, thereby justifying Bernoulli's Theorem. Afterward we shall find it possible to make use of Bernoulli's Theorem itself to extend the Multiplication Theorem to cases where the events are not independent and their probabilities are not directly obtainable.

§ 47. *The Multiplication Theorem*

We consider any two *independent* events A and B, the probabilities of which are p and p', respectively. The probability that both occur is some function of p and p'. We denote it by

$$P(A, B) = f(p, p'). \tag{45}$$

It is our problem to determine the form of the function f.

To begin with, we note that if the event A happens to include two mutually exclusive parts A_1 and A_2 (as "getting an even number with a die" includes the mutually exclusive events "getting a two," "getting a four" and "getting a six"), the compound event AB will also include two mutually exclusive parts A_1B and A_2B. By Convention II, therefore,

$$P(A, B) = P(A_1, B) + P(A_2, B),$$

and

$$p = p_1 + p_2;$$

p_1 and p_2 being, of course, the probabilities of A_1 and A_2, respectively.

With these facts before us, we need only observe that (45) is supposed to apply to *any* two independent events in order to arrive at the functional equation

$$f(p_1 + p_2, p') = f(p_1, p') + f(p_2, p'),$$

which can only be satisfied provided $f(p, p')$ is of the form $p \cdot F(p')$.[1]

Next we notice that there is no logical distinction between the occurrence of "both A and B" and the occurrence of "both B and A": the two expressions are entirely synonymous so long as the events are independent. Hence

$$f(p, p') = f(p', p),$$

or

$$p\,F(p') = p'\,F(p).$$

This requires, however, that

$$\frac{F(p)}{p} = \frac{F(p')}{p'}.$$

Since the left-hand side of this equation does not contain p', it cannot vary as the value of p' changes. But if the left-hand side does not vary with p', the right-hand side cannot. That is, $F(p')/p'$ must be a constant. Call it C. Then $F(p') = C\,p'$, and

$$f(p, p') = p\,F(p') = C\,pp'.$$

Finally, we notice that if the event A is certain to happen —that is, if $p = 1$—the occurrence of "both A and B" is

---

[1] We can readily show this if we assume the possibility of expanding $f(p, p')$ in a Taylor's Series. We write

$$f(p, p') = a + bp + cp' + dp^2 + \ldots, \tag{46}$$

whence

$$f(p_1, p') + f(p_2, p') = 2a + bp_1 + bp_2 + 2cp' + dp_1^2 + dp_2^2 + \ldots, \tag{47}$$

and

$$f(p_1 + p_2, p') = a + bp_1 + bp_2 + cp' + dp_1^2 + 2dp_1p_2 + dp_2^2 + \ldots. \tag{48}$$

If we equate coefficients of like powers in these equations we get $a = 0$ from the constant terms, $c = 0$ from the terms in p', $d = 0$ from the terms in p_1p_2, and so on. A little consideration shows, in fact, that every term of (46) which contains a power of p higher than the first will give rise in (48) to cross-products between p_1 and p_2 whereas in (47) no such cross-products can exist. It follows, therefore, that (46) must reduce to the form

$$p(b + ep' + gp'^2 + \ldots) = p\,F(p').$$

synonymous with the occurrence of "B", wherefore we conclude that $f(1, p') = Cp' = p'$. This establishes the fact that $C = 1$, and completes our proof so far as independent events are concerned.

Having thus justified the Multiplication Theorem in the case of independent events, whether or not we are able to set up the group of events required by our definition of the measure of probability, we are sure that Bernoulli's Theorem is true in general. Suppose, then that we have a pair of events A and B which are not independent. Suppose, moreover, that we make a large number M of *independent trials* of the compound event AB. Let there be, among these M trials, N_A in which A occurs and N_{AB} in which both A and B occur. Then the ratio N_A/M is not likely to differ much from $P(A)$. Similarly N_{AB}/N_A, which is the "proportion of times both A and B occur if A does," is not likely to differ much from $P_A(B)$. And finally N_{AB}/M is not likely to differ much from $P(AB)$. In fact the chance of the difference exceeding any preassigned amount can be made as small as we please in all three cases by taking M large enough.

Let us denote the differences that actually occur by δ, δ' and δ'', respectively, so that

$$P(A) = \frac{N_A}{M} - \delta,$$

$$P_A(B) = \frac{N_{AB}}{N_A} - \delta',$$

$$P(AB) = \frac{N_{AB}}{M} - \delta''.$$

Then we find by direct computation that

$$P(A)\,P_A(B) - P(AB) = -\,\delta\frac{N_{AB}}{N_A} - \delta'\frac{N_A}{M} + \delta\,\delta' + \delta''. \quad (49)$$

Now suppose $P(A)\,P_A(B) - P(AB) = d$, where $d \neq 0$. Then one or the other of the δ's must exceed $d/4$, for otherwise the right-hand side could not equal d. There is therefore

unit probability that either the one or the other will exceed $d/4$, no matter how large M may be. But this is absurd; for Bernoulli's Theorem tells us that this probability (for each one individually, and therefore also for "one or the other") is zero. Hence we must conclude that

$$P(A)\, P_A(B) = P(AB).$$

PROBLEMS

1. The fact that (44) is zero at $\eta = p$ has been said to make this the value of η for which z is a maximum. It might just as well be a minimum, however. How do we know which it is?

2. The idea of a "distribution curve" is obviously a general one. Construct a distribution curve showing the probability of runs of heads of various lengths in tossing a penny.

3. Construct a distribution function for the sums of the numbers appearing when two dice are tossed.

4. Construct a distribution function for the numbers appearing when only one die is tossed.

References for Outside Reading

1. Coolidge: *Probability*, pp. 4–10; and Chapter III.
2. Venn: *Logic of Chance*, pp. 91–95.
3. Bachelier: *Calcul des Probabilités*, Chapter II.
4. Czuber: *Wahrscheinlichkeitsrechnung*, Vol. I, pp. 128–139.

CHAPTER V

Probability and Experiment; Bayes' Theorem

§ 48. *The "Life on Mars" Paradox*

All logical processes are hedged about by a maze of fine distinctions which cannot be included in a formal symbolic expression without making it so complicated as utterly to destroy its usefulness. These distinctions are often important: so important in fact that all sorts of errors may arise through failing to remember them. In particular, in the Theory of Probability there have arisen a host of paradoxes, almost all of which are due to using *formal* logical processes in places where, upon recalling their origin, we should not expect them to apply.

One of these is a very famous one, used by the adherents of "cogent reason" to confound the "insufficient reasonists." They raise the question, What is the probability of life on Mars? Obviously, they say, we are quite ignorant. Therefore, on the basis of insufficient reason we must admit the answer to be $\frac{1}{2}$.

But the problem can also be attacked this way: What is the probability of no horses on Mars?, to which the answer is $\frac{1}{2}$; What is the probability of no cows on Mars?, to which the answer is again $\frac{1}{2}$; and this process can obviously be extended to every class of animal or vegetable. Say there are n of them. Then the probability that all these things are true: that is, that there are no horses, and no cows, and no other form of life, is $1/2^n$, which is certainly very small. The complementary probability, that there is at least *one* kind of life, is therefore near certainty. Insufficient reason has therefore given us two answers, one at least of which must be wrong.

Such is the argument as presented; but in spite of its superficial validity it is not so strong as it seems. To make the first half—the one leading to $\frac{1}{2}$ as an answer—correct, we must admit ourselves to know nothing whatever which is pertinent to answering the question. To make the second half—which leads to a little less than one—correct, we must admit, (*a*) that we know "life" to be a complex of forms, and (*b*) that the occurrence of one of these forms is quite independent of the occurrence of the rest. Of course, both postulates as to our "state of ignorance" are preposterous, which makes it just a bit harder to think straight about the problem; but we cannot object to that: the proposer of the question was entitled to postulate what he would, and we must for the time being divest ourselves of any additional knowledge we possess. Let us then put ourselves into a very unusual universe, inhabited by many kinds of life, the existence or non-existence of no one of which is prejudiced by the rest. What then is the force of the second half of the argument? Merely to demonstrate that this information *is* pertinent. To one who lived in this hypothetical universe where all the untruths upon which the second argument is based were true, the answer *would* be nearly one; just as surely as it is also the answer to the question, What is the chance of a head appearing when a handful of pennies is tossed in the air? To a six-weeks old baby, as innocent as the individual to whom the *first* half applies, $\frac{1}{2}$ would be the correct answer. To the Omniscient the answer is either unity or zero. It is neither to *us* just because we are in neither state of knowledge.

What it *is* to us I cannot say, because I am unable to evaluate the importance of all the various biological and astronomical facts which are known to bear upon the question; but I am not therefore justified in disregarding *all* I know, as is done in the first half, and calling the probability $\frac{1}{2}$, nor in disregarding *part* of what I know, as is done in the second half, and saying that it is very near certainty.

But if probability is a measure of the importance of our state of ignorance it must change its value whenever we add

new knowledge.[1] And so it does. I pick up a coin in the dark, toss it, and ask the probability that heads is uppermost. Without a doubt the answer is $\frac{1}{2}$. The coin may be biased, but I have no knowledge regarding any such bias which would make tails relatively more (or less) probable than heads. It may even be alike on both sides, but again I know nothing which makes heads more likely than tails. But suppose I now take the coin to the light and see that it really has heads on both sides. Immediately the chance of heads being uppermost if I toss it again (or that heads were uppermost when I tossed it before) changes to 1.

The knowledge that the penny is alike on both sides is of such an absolute character that it enables us to state at once the change that it makes in our probability. We might, however, have gotten *inferential* knowledge instead. For example, we might have tossed it a large number of times —100, say — and found that heads appeared every time. Such a thing might happen either with a perfectly good penny or with one that has heads on both sides. It might also happen with a penny which was badly biased. We cannot say with assurance just which situation exists, but certainly the probability that heads will appear upon the next toss is no longer $\frac{1}{2}$. *It has been changed by experiment.*

This, then leads us to the question of *how much* it has changed, to which Bayes' Theorem is the clue.

§ 49. *Bayes' Theorem*

We return to the consideration of the equations (20) and (21). Obviously, (20) can be written either in the form

$$P(AB) = P(A)\, P_A(B) \qquad (20)$$

or in the form

$$P(AB) = P(B)\, P_B(A), \qquad (50)$$

[1] It is sometimes objected that this makes probability a "personal" matter. So it does, in a sense, but only in the sense in which it *is* a personal matter — that is, in the sense in which it depends upon individual differences of knowledge. No one would say the probability of heads appearing when a penny is tossed was $\frac{1}{2}$ to one who *knew*. It is only that for all of *us* when none of us knows a thing about it.

provided that, when any questions of order of occurrence are involved, the same order shall apply to both.[1]

We recall that our symbols have the following meanings:[2]

$P(AB)$. The probability that A happened and was followed by B.

$P(A)$. Nothing is known about whether or not B happened. $P(A)$ is the probability that A did.

$P(B)$. Nothing is known about whether or not A happened. $P(B)$ is the probability that B did.

$P_A(B)$. A is known to have happened. This is the probability that it was followed by B.

$P_B(A)$. B is known to have happened. This is the probability that it was preceded by A.

Since $P(AB)$ means exactly the same thing in both equations, we may equate their right-hand members and solve for $P_B(A)$. The result is

$$P_B(A) = \frac{P(A)\ P_A(B)}{P(B)}. \tag{53}$$

---

[1] In the footnote in § 22 we called attention to the fact that the chance that "a man is shot and dies," taken in that temporal, or causative, sequence, is *not* the same as the chance that "he dies and is shot," taken in that order. Suppose we call being shot the "event A," and dying "event B." Then the *first* order is AB, the other BA, and their respective probabilities may be denoted by $P(AB)$ and $P(BA)$. To the first of these corresponds the pair of equations (20) and (50); and to the second an exactly similar set which we may write in the form

$$P(BA) = P(B)\ P_B(A). \tag{51}$$

$$P(BA) = P(A)\ P_A(B). \tag{52}$$

Obviously the right-hand members of these equations are symbolically equivalent to (20) and (50). But the equivalence stops with the symbolism. For $P(B)$ in (50) means "the probability that a man dies," and the $P_B(A)$ means "the probability that a dying man *has been* shot"; while in (51), though the $P(B)$ means just what it did before, the $P_B(A)$ means "the probability that, having died, he *will be* shot."

We shall interest ourselves only in (20) and (50), though the other pair also lead to a sort of "Bayes' Theorem" which, under certain circumstances, has a slightly different meaning from the usual one.

[2] The use of the past tense is of no significance: present or future would do just exactly as well, except that grammatically the perfect and past are slightly less clumsy than future perfect and future, while the "tenseless present" of logic destroys the sharpness of the distinction between past and present.

In other words, knowing the probability $P(A)$ which would apply if we knew nothing of whether B occurred or not, and the probability $P(B)$ which would apply if we were in total ignorance regarding A, and also the probability of B following A: knowing these, and also knowing that B *did* occur, we can find the probability that it was preceded by A. *Equation* (53) *is therefore the clue to interpreting the influence of new knowledge upon probability. It is called Bayes' Theorem.*

Before illustrating its use, we may note that it is usually given in the somewhat different form to which it reduces when (21) is substituted for the denominator of (53). This form is

$$P_B(A) = \frac{P(A)\, P_A(B)}{\sum_A P(A)\, P_A(B)}, \qquad (54)$$

or, in words:

BAYES' THEOREM: *If the event B never occurs unless preceded by one or the other of a set of events A, the unconditional probabilities of which are P(A), and if B is known to have happened, the chance that it was preceded by a particular one of the events A is a fraction, the numerator of which is the product of the probability of this particular event by the chance of it being followed by B, while the denominator is a sum of exactly similar terms, one for each of the events A.*[1]

§ 50. *Some Instructive Illustrations; Bertrand's "Box Paradox"*

The following example is an illustration in which there can be no question as to our possession of the information necessary to the use of Bayes' Theorem:

EXAMPLE 34.—*Three boxes have in them two coins each. In one box both are gold, in one both are silver, in the other they are mixed.*

[1] It must be noticed that throughout the entire formula every symbol refers to the sequence "A followed by B." From (51), (52) and (21) an equation can be obtained which is identical in form, but in which the sequence is always "B followed by A." Hence (53) and (54) are true when consistently read either way; but when the order is of any consequence, as it sometimes is, care must be taken not to mix the two.

Outside they are of identical appearance. A man chooses a box and takes out a coin, which proves to be gold. What is the chance that the other coin in the box is also gold?

It is not necessary to use Bayes' Theorem to solve this problem. The solution can be obtained directly as follows: Each of the three gold coins is as likely to be the one chosen as the other. Two are in the box in which " the other is also gold." Hence the answer is $\frac{2}{3}$. It will, however, serve our purpose best to get the answer in two rather roundabout ways corresponding to (53) and (54), respectively, and for this purpose we choose the following symbolism: A_{gg} for the event " chooses the box with two gold coins," and A_{gs}, A_{ss} for the alternative possibilities; B_g for " picks up a gold coin," and B_s for the opposite.

If we know nothing about what has transpired,[1] the probability of A_{gg} is $P(A_{gg}) = \frac{1}{3}$. Likewise, if we know nothing about what has transpired, one coin is as likely to have been picked up as another. Hence $P(B_g) = \frac{1}{2}$. But *if* the box with two gold coins was chosen, the chance of picking out a gold coin was $P_{A_{gg}}(B_g) = 1$. Hence (53) gives for the probability, after a gold coin has been seen, of the box having two gold coins,

$$P_{B_g}(A_{gg}) = \frac{P(A_{gg})\, P_{A_{gg}}(B_g)}{P(B_g)} = \frac{\frac{1}{3}\cdot 1}{\frac{1}{2}} = \frac{2}{3}.$$

Or we can phrase the solution this way: $P(A_{ss})$ and $P(A_{gs})$ are also $\frac{1}{3}$, while the conditional probabilities of getting a gold coin *if* the chosen box is *ss* or *gs*, respectively, are $P_{A_{ss}}(B_g) = 0$ and $P_{A_{gs}}(B_g) = \frac{1}{2}$. Hence (54) becomes:

$$P_{B_g}(A_{gg}) = \frac{P(A_{gg})\, P_{A_{gg}}(B_g)}{P(A_{gg})\, P_{A_{gg}}(B_g) + P(A_{ss})\, P_{A_{ss}}(B_g) + P(A_{gs})\, P_{A_{gs}}(B_g)}$$

$$= \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0} = \frac{2}{3}.$$

As was to be expected, the answer is the same in all three cases.

[1] Or, " will transpire."

§ 51. *Some Instructive Illustrations; Another Urn Problem*

The following is another problem to which Bayes' Theorem affords the solution:

EXAMPLE 35.—*A box has had ten balls put in it by the following procedure: An auxiliary container holds equal numbers of black and white balls, thoroughly mixed. A blindfolded man picks one out and places it in the box. He is watched by an assistant who immediately puts another ball* of the same kind *in the auxiliary container and stirs up the contents. Then a second ball is drawn, and so on. After the box has received its ten balls, an experiment is performed by drawing balls repeatedly from it, noting the color, and replacing them. Five such attempts show four black balls and one white. What is the most probable contents of the box?*

Before the experiment, the probability of A white and $10 - A$ black balls was obviously

$$P(A) = C_A^{10} \left(\tfrac{1}{2}\right)^{10}.$$

If A white balls are there, the chance of drawing one white and four black in five attempts (this is the " event B " which is known to have happened), is

$$P_A(B) = C_1^5 \left(\frac{A}{10}\right)\left(\frac{10 - A}{10}\right)^4.$$

Hence Bayes' Theorem gives

$$P_B(A) = \frac{\left(\tfrac{1}{2}\right)^{10} C_A^{10} C_1^5 \left(\dfrac{A}{10}\right)\left(\dfrac{10 - A}{10}\right)^4}{\sum_{A=0}^{10} \left(\tfrac{1}{2}\right)^{10} C_A^{10} C_1^5 \left(\dfrac{A}{10}\right)\left(\dfrac{10 - A}{10}\right)^4}.$$

Obviously the $\left(\tfrac{1}{2}\right)^{10}$, C_1^5, and the 10's by which the A's are divided are common factors of all terms. Hence they can be cancelled, thus yielding

$$P_B(A) = \frac{C_A^{10} A (10 - A)^4}{\sum_{A=0}^{10} C_A^{10} A (10 - A)^4}. \tag{55}$$

Actual computation gives the results shown in Table XII.

The required probabilities are in the last column. Obviously 4 white and 6 black is the most likely composition.

Before the experiment was performed the most probable composition was half and half. The preponderance of black results has altered the probabilities in the direction which common sense would dictate; but the experimental proportion ($\frac{1}{5}$ white) is still only one-third as likely as the most probable proportion ($\frac{2}{5}$); and only half as likely as the half-and-half division.

TABLE XII

A	C_A^{10}	$A(10-A)^4$	$C_A^{10}A(10-A)^4$	$P_B(A)$
0	1	0	0	0.00000
1	10	6561	65,610	0.01837
2	45	8192	368,640	0.10323
3	120	7203	864,360	0.24204
4	210	5184	1,088,640	0.30484
5	252	3125	787,500	0.22051
6	210	1536	322,560	0.09032
7	120	567	68,040	0.01905
8	45	128	5,760	0.00161
9	10	9	90	0.00003
10	1	0	0	0.00000
			$\Sigma = 3,571,200$	

Suppose, instead of five trials, fifty were made, giving 10 white and 40 black. Common sense tells us the experimental evidence must now be given more weight than before, but common sense does not tell us *how much more.* Bayes' Theorem does. For now we must change the powers of A and $(10 - A)$ in (55), to 10 and 40 instead of 1 and 4. This, however, replaces all the numbers in the third column of Table XII by their *tenth* powers, which quite obviously causes the third entry (8192) to increase very much more than any of the rest. In fact, the next largest entry (7203) is only about 0.9 times as large as 8192; after the tenth powers are taken it will be about $\frac{1}{3}$ as large. Hence when these numbers are multiplied by 45 and 120, respectively, the latter of which is about three times the former, they will yield new probabilities

which do not differ much in magnitude. Of course the rest of the new probabilities will be very much smaller.

The exact results are shown in the second column of Table XIII; while the third column contains the probabilities after 500 tests have given 400 black and 100 white balls. In the last case the experimental evidence completely outweighs any preconceptions arising from our knowledge of how the box was filled. It is *99.999 per cent sure* that there are just two white balls.

TABLE XIII

A	$P_B(A)$	
	After 50 Tests	After 500 Tests
0	0.00000	0.00000
1	0.01334	0.00000
2	0.55276	0.99999
3	0.40714	0.00001
4	0.02657	0.00000
5	0.00020	
6	0.00000	

§ 52. *Some Instructive Illustrations; The Bad Penny*

Let us consider one more example of the use of Bayes' Theorem—again in connection with a problem of no practical consequence, though it may aid us in gaining an accurate picture of what the theorem is good for.

We suppose, to begin with, that a "penny" is given us, and that we are asked to determine whether it is alike on both sides, or normal. Of course the sensible thing to do would be to look and see; but this is not a sensible problem. Instead we propose to find out by tossing it repeatedly and noting what shows up. At the start, before any experiments have been made, there is a certain probability that it is bad (alike on both sides). We cannot say what this is: we may have been told by someone in whom we have considerable confidence that it *is* bad, and in that case the probability is high. Or

we may have been told that it is good, and the probability is low. Whatever it is, let it be denoted by $P_0(b)$, and the complementary probability that the penny is good by $P_0(g)$.

Suppose, now, that n throws are made and all of them result in heads. If the penny has heads on both sides, the chance of this is $P_b(n) = 1$; while if it is a good penny the chance of a run of n heads is only $P_g(n) = (\frac{1}{2})^n$. Substituting these values in (54) we obtain at once, as the probability, *after* the experiment, that the penny is bad,

$$P_n(b) = \frac{P_0(b) \cdot 1}{P_0(b) \cdot 1 + P_0(g) \cdot (\frac{1}{2})^n},$$

or, if we write $P_0(g)/P_0(b) = k$, for simplicity,

$$P_n(b) = \frac{1}{1 + \frac{k}{2^n}}.$$

This result depends upon two things, as it should: upon the degree of our assurance before we experiment, and upon the number of trials carried out. If there is no reason to suspect the penny—for instance, if it is a coin casually picked up on the street—$P_0(g)$ certainly exceeds $P_0(b)$ and k is large; for there is obviously much greater likelihood of happening on a good coin than a bad one. Suppose we assume that there is only one chance in a million that it is bad, which makes $k = 1{,}000{,}000$. Then after ten heads have appeared without interruption the chance of it being bad is[1] $P_{10}(b) = \frac{1}{1000}$. The new probability, though still small, is very much larger than before the experiment was performed. If the run continues and twenty heads appear, the probability becomes $\frac{1}{2}$: twenty heads, in other words, just counterbalance our preconceived notion that the penny was very probably good. Thirty heads, on the other hand, give a probability $P_{30}(b) = 0.999$. There is now only one chance in a thousand that the penny is not bad.

[1] These are all round numbers based upon the approximation $2^{10} = 1000$, which is plenty near enough for our purpose here. The true value is 1024.

Of course these figures have resulted only because of our assumption of 1 : 1,000,000 as the *a priori* chance that the thing is bad, and we must not lose sight of the fact that we have no way to check this guess. As exact values they are worthless, but they serve to illustrate how rapidly an uninterrupted run of luck may wipe out a strong presumption in one direction and replace it by an equally strong presumption the other way. Had we chosen any other number than a million the result would have been much the same: if we had guessed the chance of the penny being bad to be the inconceivably small number 0.000,000,000,001, it would still require only fifty heads in succession to replace this probability by 0.999.

Finally, suppose we were mathematically certain that the penny was good—suppose, in other words, that $P_0(b)$ were zero, not approximately, but absolutely. Then k would be infinite, and so also would $k/2^n$ no matter how great n might be. In this case $P_n(b)$ would be zero for every value of n. This, too, is as it should be, for experimental evidence is trivial beside infallible certainty. How we could ever reach such a state of absolute assurance I do not know; but if we could, no amount of experimentation should be allowed to shake our faith.

§ 53. *The Uses to Which Bayes' Theorem May Be Put*

There are many important problems of a scientific character which are essentially similar to the one we have just been considering. Almost any instance where a scientific generalization is to be made from a limited amount of data could be cited in illustration: for example, the conclusion that all electrons have like charges. One bit of evidence consists of the fact that a certain number have been isolated and their charges measured, and there is a considerable volume of indirect evidence. But how sure are we of the rule? Not absolutely sure, certainly; for though it is pretty generally believed, it has occasionally been contested.

We would be glad, if we could, to get a measure of our certainty in cases such as this; but there is nothing in the

Theory of Probability to aid us except Bayes' Theorem, which we can seldom use for the purpose because, as in the problem of the last section, we cannot measure the unconditional probabilities. Any of the answers which we have obtained would be correct IF k had the values assumed; other answers could easily be obtained IF k were something else; but as long as we do not know what k is, we cannot be certain of any of them.

To such problems, then, no *exact* answer can be expected from the use of Bayes' Theorem; not because of any logical uncertainty as to the theorem itself,[1] but because we do not possess the data necessary for its use. On the other hand it is often of service in dealing with problems to which *qualitative* answers are acceptable. Thus in the case of the bad penny example of § 52, we might desire to know how many throws would be required to justify the belief that the penny is bad. To this question the answer "not less than twenty nor more than fifty" can safely be given—not very close limits, to be sure, but at least indicating the order of magnitude. A smaller number of throws would give us very little information, because of the inherent probability that the penny is good; and a larger number would not increase our certainty to any material extent.[2]

There are also cases in which, although we have no exact means of determining the *a priori* probabilities, we have reason to believe we know them to a fair degree of approxima-

[1] Due to numerous inexact statements which have been made of it, Bayes' Theorem has been the subject of much adverse criticism, and some authorities have even gone so far as to reject it entirely. At present, however, this criticism seems to be dying out, the commonly accepted view being much the same as that stated above: that it is just as sound logically as any other part of the Theory of Probability, and may be trusted to give reliable results *when we can get a grip on it.* The trouble is that we so seldom can.

[2] This is true, not only because the result given by Bayes' Theorem is already substantially 1, but also because other alternatives, originally neglected because of their inherent absurdity, become of greater and greater consequence as the run proceeds. Suppose, after a run of 100 heads, a tail were to appear. The chance of this happening with a good penny is so extremely small, that hallucination or substitution would merit serious consideration.

tion. An example is given in the section immediately following. To such problems I think it is wise to apply the theorem, for while no *exact* information can be expected from this course, *better* information will be obtained than can be gotten in any other way. It will need judicious interpreting, of course, but that is not an uncommon state of affairs when mathematical reasoning is applied to scientific problems.

Finally, Bayes' Theorem can often be applied with absolute rigor in the course of a formal mathematical argument. It is unnecessary to introduce artificial illustrations of its use in this direction, however, as enough examples will arise naturally in the course of our further studies.

§ 54. *Some Instructive Illustrations; An Elementary Problem in Sampling*

Consider the following problem in the testing of factory output:

EXAMPLE 36.—*A factory produces a certain type of screw as a standard product. The screws are collected at the machine in boxes of 1200 each. Long experience has shown that the proportion of these boxes which contain various percentages of bad screws is substantially as follows:*

Per Cent of Bad Screws in the Box	Proportion of Boxes Observed to Contain this Percentage of Bad Screws
0	0.78
1	0.17
2	0.034
3	0.009
4	0.005
5	0.002
6	0.000

Two per cent badness has been adopted as a manufacturing standard; that is, any box which contains 2 per cent or less of bad screws is regarded as satisfactory, the aim of the inspection process being to reject those which are poorer. The normal inspection consists in the examination of 50 screws out of each box. A particular box, produced at a time when

there was no special reason to suspect that the machines were not operating properly, showed 6 bad screws under normal inspection. What is the probability that the manufacturing standard had not been maintained in the production of this box?

We know from Bernoulli's Theorem that the *proportions* listed in the second column of this table are probably good approximations to the *probabilities* of the various percentages of badness. They may therefore be used as the values of $P(A)$ in (54), the "event A" standing successively for the various possible percentages of bad screws in the box. The conditional probabilities then follow at once

TABLE XIV

A	$P(A)$	C_6^{12A}	$\frac{(1200-12A)!}{(1200-12A-44)!}$		$P_B(A)$
0	0.78	0.0000	13.721^{134}	0.000^{139}	0.000
1	0.17	9.2400^{2}	8.746	0.014	0.004
2	0.034	1.3460^{5}	5.548	0.254	0.070
3	0.009	1.9478^{6}	3.503	0.614	0.170
4	0.005	1.2272^{7}	2.201	1.351	0.374
5	0.002	5.0060^{7}	1.376	1.378	0.382
6	0.000				0.000
				3.611^{139}	

from (25), p and q being, respectively, 6 and 44, while m and n are $12A$ and $1200 - 12A$. When these are substituted in (54), and obvious common factors are cancelled out, they yield the formula

$$P_B(A) = \frac{P(A)\, C_6^{12A}\, \dfrac{(1200-12A)!}{(1200-12A-44)!}}{\sum_A P(A)\, C_6^{12A}\, \dfrac{(1200-12A)!}{(1200-12A-44)!}}.$$

It is a comparatively simple matter to compute the values of this expression. The outline of the computation is shown in Table XIV. The second column contains the values of $P(A)$ as given in the statement of the example; the third column contains the binomial coefficients as taken from Appendix III; while the fourth column contains the values of $\dfrac{(1200-12A)!}{(1200-12A-44)!}$. These were found by the use of Appendix II, in which the logarithms of large factorials are given. When these three columns are multiplied together they give the fifth. Each of the entries in this column is the numerator of the fraction

which represents $P_B(A)$ for the corresponding value of A; and the sum of the entire column, 3.611^{139}, is the denominator for all of them. Hence the desired probabilities are to be found by dividing every entry by 3.611^{139}. The results are shown in the last column.

The chance that more than 2 per cent are bad is the sum of the last four entries in this column, and is so large as to render it highly probable that trouble exists.

It is unnecessary to say that if the manufacturing situation postulated in the example really existed, this sort of computation would be made once for all for such results as were likely to be met. Thereafter it would be necessary only to refer to the tabulated probabilities to learn the significance of any set of results; or, more probably, that number of defective screws would be determined for which there was an even chance of trouble existing, and some routine would be established to assure that the trouble was quickly located and corrected. The exact manufacturing conditions would determine what sort of routine was best.

However, the entire problem is, in a way, idealistic. In the first place, we have tacitly assumed in its statement, and in computing our results, that such proportions of product as 1 per cent, 2 per cent, etc., might be bad, but not 1.5 per cent or other fractional percentages. Obviously, this is not the case: for out of the 1200 screws in a box any multiple of $\frac{1}{12}$ per cent is a possibility; and in a more general type of problem the variable might be capable of taking almost any values. We must therefore interpret 0 per cent as including all cases for which the actual percentage is less than $\frac{1}{2}$ per cent; 1 per cent as including all other cases less than 1.5 per cent, and so on. It is a matter of experience that such grouping of data frequently causes so little error that the added cost of more complete computation is not warranted. Hence, on the whole, the computation is probably as good as any practical situation of its kind is likely ever to warrant.

PROBLEMS

1. The following has been given as an example of the incorrect use of Bayes' Theorem when applied to Example 34. (It deals with the chance of getting unlike coins instead of like ones.)

"There is a $\frac{1}{2}$ chance that the coin first seen shall be gold. When gold has been seen, we know that we have chosen one of the first two boxes, but we do not know which. They are equally likely, hence the chance for a gold coin followed by silver is $\frac{1}{4}$. There is an equal chance for a silver coin followed by gold. Hence the total chance is $\frac{1}{2}$."

What is wrong with this argument?

2. The same author gives the following as a correct solution:

"If a gold coin has been seen, the *a priori* chance for the first or second box is $\frac{1}{2}$, but whereas the first has a chance 1 of showing a gold coin the first time, the second has only a chance $\frac{1}{2}$ of doing so. The probability that the gold coin is in the second box is

$$\frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot\frac{1}{2}} = \frac{1}{3}.$$

and there is a similar probability for a silver coin."

This argument is also wrong. Explain why.

3. A box has been filled as in Example 35. We are told, and have complete faith in the information, that 50 balls have been drawn and have given 10 white and 40 black. We, however, make an independent test by drawing 5 balls, all of which turn out white. What effect has this upon the probabilities of various proportions of the two colors?

4. If the proportions of Example 35 turned out to be half and half, how would the probabilities be affected?

5. In the "bad penny" problem of § 52, suppose the first six throws show a run of five heads followed by a tail. What happens to the probabilities?

6. Suppose instead that 600 throws showed a run of 599 heads followed by a tail. Would the situation be any different?

(More than a simple "yes" or "no" is wanted to this problem. Explain as clearly as you can your reactions to it as a problem in logic.)

References for Outside Reading

1. Coolidge: *Probability*, Chapter VI.
2. Arne Fisher: *Mathematical Theory of Probabilities*, pp. 54–67.
3. Venn: *Logic of Chance*, pp. 179–202, 249–264.
4. Czuber: *Wahrscheinlichkeitsrechnung*, pp. 189–210.
5. Bachelier: *Calcul des Probabilités*, Chapter XXIII.
6. Laplace: *Théorie Analytique des Probabilités*, Chapter VI.
7. Molina and Crowell: *Deviation of Random Samples from Average Conditions and Significance to Traffic Men*, Bell System Technical Journal, Vol. 3 (1924), pp. 88–99.

CHAPTER VI

Distribution Functions and Continuous Variables

§ 55. *The Random Choice of a Point on a Line Segment*

So far in our study we have dealt only with numbers and events which were essentially discrete. For instance, if a penny is tossed, two discrete events—heads and tails—are possible. In the nature of things there can be no intermediate state of affairs. Or if the penny is tossed repeatedly we may get various numbers of heads: but always integral numbers, never such a result as 1.732 heads.

There are, however, many sets of events which do not fall into such discrete catagories, although they are proper subject matter for a mathematical study of probability. For example, in the production of lamps the ideal situation would be to have them all of exactly the same resistance. The factory process aims at that, but certainly does not accomplish it. Neither is it true that a certain discrete set of resistances may occur, and that others are totally impossible. Instead, resistance can vary continuously.

This difficulty presents itself immediately: if there are not *discrete events*, there cannot be groups and sub-groups suitable for the application of our definition of probability. It would seem, therefore, that some entirely new definition would be necessary. This is not quite true, however, for the difficulty really resides in the notion of a continuous variable rather than in the definition of probability, and may be removed by exactly the same sort of argument as is used in defining irrational numbers.

It is the principal purpose of the next few sections to make this clear. As the first step in this direction we begin with an example of so special a form that it may appear at first sight to have little to do with the fundamental problem, but from it, by successive generalizations, we shall be able to obtain what we seek.

EXAMPLE 37.—*The Random Choice of a Point on a Line Segment: The perimeter of a well-balanced wheel is of unit length, and carries a uniform scale, such as the scale of a yard stick. A pointer of negligible thickness is set up opposite this wheel, and the wheel is spun. When it comes to rest the pointer indicates some number on the scale. What is the chance that it lies in the interval between two numbers a and b?*[1]

To start with, suppose this interval is 0.7 to 0.8. We know that the ten segments marked off by the points 0.0, 0.1, 0.2, . . . , 0.9 are all of equal length. There are therefore ten equally likely and mutually exclusive events corresponding to the ten segments within which the pointer may possibly rest. Of these only one belongs to the desired subgroup. Here our original definition applies, and we find that the probability of the pointer resting between 0.7 and 0.8 is $\frac{1}{10}$.

If we modify the requirements of the problem by demanding the probability that the pointer come to rest between 0.70 and 0.71, we are able to divide up the entire perimeter of the wheel into 100 equally likely divisions; and we conclude at once that the desired probability is $\frac{1}{100}$.

To take a somewhat more general case, let us call the length of the interval $b - a = x$. If x is a rational number—that is, if it can be represented as the quotient of two integers n/m—it is possible to start from a and lay off the perimeter of the wheel into m equal divisions, of which exactly n are

[1] It may avoid certain logical difficulties to regard this interval as containing a but not b. That is, if the pointer rested on a it would be said to be in the interval, but if it rested on b to be without it. The only purpose of such a convention is so to arrange matters that the sum of the intervals from a to b and from b to c is just exactly the interval from a to c. If we made any other convention, we would either include one point twice, or omit it altogether.

included in the interval from a to b. Obviously the probability of the pointer coming to rest within this interval is just $n/m = x$: that is, it is the length of the interval. *Hence, whenever the length of the interval is a rational number, the probability of coming to rest in it is equal to its length.*

But what if the interval is not representable as the quotient of two integers? In this case our fundamental definition of the measure of probability breaks down; for it is impossible to find any method of division which divides both the perimeter of the wheel and the segment exactly.

Since this difficulty arises only when the number x is irrational, it is natural to go for relief to the branch of mathematics which deals with the nature of irrational numbers. When we do this, we find an even more fundamental difficulty: an irrational number cannot be written in our ordinary number system at all, and requires a totally different set of ideas for its definition.[1] We get our first insight into the relation which these irrational numbers bear to the rest of our number system —and at the same time an indication of how to overcome our present difficulty — by considering how we deal with them in practical life.

Suppose we choose $1/\sqrt{2}$ as our illustration. We ordinarily write it as 0.7 or 0.707, or 0.7071; understanding in each case that what we have written is the *nearest* tenth, or thousandth, or ten-thousandth. We could easily frame the

---

[1] Some rational numbers cannot be written as "decimal fractions" because the base of our number system is 10. The fraction $\frac{1}{3}$ is of this class, for it leads to the "repeating decimal" 0.333 . . . ; but if the basis of our number system were a multiple of 3 this difficulty would disappear. Thus, if the base were 12 the fraction $\frac{1}{3}$ would be represented by the decimal 0.4, for 0.4 would then mean $\frac{4}{12}$ instead of $\frac{4}{10}$.

The difficulty which we have above is not of this type, but is a fundamental fact in the logic of number and persists whatever base may be chosen. It can be shown that in all cases where the difficulty is due to the choice of the base 10, the succession of digits which makes up the decimal ultimately begins to repeat itself; that is, that all rational numbers can be represented by either "repeating" or "terminating decimals." Irrational numbers, on the other hand, not only cannot be represented by terminating decimals nor by the quotient of two integers: they cannot even be represented by repeating decimals.

same idea in the form of a set of inequalities if we so desired. For example:

$$0 < \frac{1}{\sqrt{2}} < 1,$$

$$0.7 < \frac{1}{\sqrt{2}} < 0.8,$$

$$0.70 < \frac{1}{\sqrt{2}} < 0.71,$$

$$0.707 < \frac{1}{\sqrt{2}} < 0.708.$$

Such a set of inequalities could be extended as far as we pleased,

This is not only a *practical* expedient, however. *Logically* also, it leads to important consequences; for it shows that we can set up a sequence of *rational* numbers, 0, 0.7, 0.70, 0.707, . . . , approaching $1/\sqrt{2}$ as a limit, though each number is known to be less than $1/\sqrt{2}$; and another sequence of *rational* numbers, 1, 0.8, 0.71, 0.708, . . . , which likewise approaches the limit $1/\sqrt{2}$, though each of its terms is greater than $1/\sqrt{2}$. The same is true of any other irrational number: π, for example, is approached by the sequence 3, 3.1, 3.14, 3.141, . . . , every term of which is less than π; and also by the sequence 4, 3.2, 3.15, 3.142, . . . , every term of which exceeds π. We conclude, therefore, that *every irrational number is the limit of at least two sequences of rational numbers, one approaching it from below, the other from above.*

Let us now return to the discussion of our revolving wheel. If x is irrational, we know that there must be a sequence of rational numbers

$$\frac{n_1}{m_1}, \frac{n_2}{m_2}, \frac{n_3}{m_3}, \ldots,$$

all *less* than x, but approaching it as a limit; and another sequence

$$\frac{N_1}{M_1}, \frac{N_2}{M_2}, \frac{N_3}{M_3}, \ldots,$$

all greater than x, but approaching it from above. Suppose that we locate on the perimeter of the wheel the four points a, $a + n_i/m_i$, $a + x$, $a + N_i/M_i$; the subscript i corresponding to some term in the sequence.

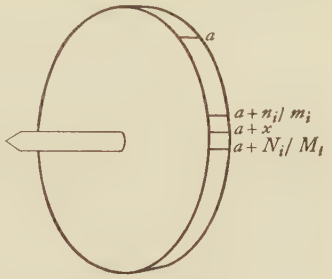


FIG. 11.

We know that the probability of the pointer resting in the interval a to $a + n_i/m_i$ (see Fig. 11) is equal to n_i/m_i. But it cannot rest in this interval without also being in the interval between a and $a + x$. Therefore, $P(x)$ is *at least as great as* n_i/m_i, $P(x)$ being the desired probability. Similarly, whenever the pointer is within the interval $(a, a + x)$ it is also within the interval $(a, a + N_i/M_i)$. Hence $P(x)$ *cannot exceed* N_i/M_i. That is:

$$\frac{n_i}{m_i} \leqq P(x) \leqq \frac{N_i}{M_i}.$$

This inequality is true for any value of i. As i gets larger and larger, however, the two quantities n_i/m_i and N_i/M_i converge to the same limit x, and as $P(x)$ constantly lies between them, no matter how close to x they may come, it follows of necessity that $P(x) = x$.

The probability of the pointer lying between a and b is equal to the length of the interval regardless of whether that length is rational or not.

We have required that the perimeter of the wheel be of unit length. This, however, instead of defining the size of the wheel, merely defines the unit in terms of which all lengths are measured. The theorem only states, therefore, that the probability of lying within any interval is equal to the length of that interval *measured in terms of the perimeter of the wheel as a unit of measure.* As the *probability* would be the same no matter what unit of length was used, we conclude that:

The probability of lying within an interval of length x upon a wheel the circumference of which is L, is x/L.

This is, in a sense, an extension upon the definition of "the measure of probability," for that definition cannot be applied to this problem at all. It is made necessary by the fact that "length" is a continuous variable. But the extension is a theorem, rather than a new axiom, for it is a logical consequence of the accepted relationship of irrational to rational numbers.

Finally, we note that the revolving wheel of our example serves no other purpose than to assure that the chosen point is equally as likely to lie within one interval as another. The same probability would have been obtained with any other mechanism for which equal intervals were equally likely. We can therefore take as the final form of the theorem at which we have arrived:

If a point is placed upon a line segment of length L in such a way that equal intervals are equally likely to contain it, the chance that it lies within an interval of length x is x/L.

Such a point will be said to be placed "at random" on the line.

§ 56. *A Paradox Associated with the Random Choice of a Point on a Line Segment*

There is a curious paradox associated with this matter of choosing a point on a line segment which is of interest because it throws some light upon — or at least calls attention to — the meaning of zero probability.

The paradox arises in connection with the question, What is the probability that the point chosen is the midpoint of the line L? If we construct an interval of length x about this midpoint, we know that the chance of the point lying within this interval is x/L. It is possible, however, for the point to lie within the interval and still not coincide with the midpoint: we conclude therefore that the desired probability is less than x/L. This must be true no matter how small the interval becomes, which is only possible provided the desired probability is zero. *That is, the probability that a point placed at random upon a line bisects that line is zero.*

This argument has been carried out for the midpoint of the line L. It is obvious, however, that it could have been carried out equally well for any other point.[1] We therefore conclude that *the probability that a point placed at random upon the line L coincides with any pre-assigned point, is zero.*

However, the random point must have fallen somewhere on the line, that is, it must have fallen upon some point or other. But according to the statement in the last paragraph the *a priori* probability that it would fall where it did was zero. If, then, zero probability means that the event cannot occur, the random point has done the impossible, for it cannot be where it is, and yet it most assuredly is there.

The paradox is only one of many associated with the concept of infinitely large numbers, or of limiting processes in general.[2] To understand it thoroughly we must go back to our original definition of probability, according to which we deal with a group of m equally likely events, and ask for the probability that one of a sub-group n occurs. The measure of the probability is the ratio n/m. So long as m is fixed this ratio can only vanish provided n is zero; that is, provided none of the possible events is included in our sub-group.

In such cases, zero probability means impossibility. If, however, by some means we keep the number n constant and increase m indefinitely we can cause the probability of success to be as small as we please. For example, if there are ten balls in a box, n of which are red and $10 - n$ white, the chance that a ball drawn from this box is red is $n/10$. It can only be zero provided there are no red balls in the box, in which case the drawing of a red ball is obviously impossible. The same is true if the number of balls is a hundred, a thousand, or a million. But if, having started with ten balls in the box, we add larger and larger quantities of white balls, the chance of drawing a red ball from the box obviously becomes smaller and smaller, and can in fact be made as small as we please by adding enough white balls. Thus, if there was originally only one red ball, so that the original probability was $1/10$, we are able to make this probability less than $1/100$ by adding 100 white balls (it is then

[1] We are prone to feel — though our better judgment tells us it is not true — that it would be more unusual for the point to bisect the interval than to fall on some other point that "had nothing peculiar about it." Symmetrical effects always impress us unduly. There is a case on record where each of four bridge players drew a complete suit. "Most remarkable," we say: yet actually no less probable than the last hand you held. Each had a probability of $4.48 \cdot 10^{-28}$. The remarkable thing about it was its symmetry, not its rarity.

[2] It must be remembered that infinity is not a number, but a characterization of the behavior of a variable which ultimately exceeds any number.

1/110). If we are asked to make it less than one in a million we can do it by adding a million white balls. In general, if we are asked to make it less than any number ϵ we can do this by adding m balls, where $m > 1/\epsilon$. So long as m is finite the resulting probability is also finite and we have no dilemma. We are merely assured of the fact that the probability is exceedingly small. But as m becomes infinite the probability in question becomes zero. Logically, therefore, zero probability means impossibility only when the group of events is finite in extent; when the group of events is infinite it means merely that the sub-group of favorable events is but an infinitesimal part of the whole.[1]

There is a certain sense in which theory and practice diverge at this point, for many things which are logically possible are practical impossibilities. Looked at from a practical standpoint, zero probability always means impossibility. With an infinity of balls the red ball might *conceivably* be drawn but the drawing of it would be a miracle rather than a "practical possibility."[2]

Now the placing of a point upon a line is in exactly this class. We could not accurately bisect the line if we tried our utmost to do it, much less if we cut it at random; and yet the bisecting of the line is not a logical impossibility.

§ 57. *Extension of the Significance of the Preceding Paragraphs*

So far in this chapter we have talked continually about "placing a point at random" on a line segment. In the first instance we thought of this as being done by means of a machine of a special type. Then we made our concept somewhat more general by dropping the conception of the machine entirely, and talking merely about the placing of the point, regardless of the mechanism employed. It has now become desirable to recognize the fact that the arguments which we have been carrying out have an even more general significance than this, and that the talk about "points" and "line segments" has been merely a form of expression which recom-

[1] It is unfortunate, in a way, that mathematics has no separate notation for numbers reached through limiting processes. If it had, we could say that zero (the number) as a probability means the thing is impossible; while zero (the limit) must be interpreted in terms of the limiting process which gives rise to it.

[2] There are propositions which are *logically* as well as *practically* impossible. We usually call them absurdities. For example: "x is less than a number which is less than x"; or, "He is his mother's father."

mended itself by making it somewhat easier to visualize the steps of the argument.

To place a point upon a line which carries a scale is equivalent to choosing a number. Conversely, whenever any quantity whatever is measured, its magnitude can be plotted, thereby determining a point. Hence *locating points on lines* and *measuring quantities* are interchangeable ideas: whatever is true of one of them is true *mutatis mutandis* of the other also. The two fundamental results which have so far been obtained are therefore capable of being framed as follows:

If by any process whatever a number is obtained concerning which two things are known: (1) that it cannot be less than x_1 *nor greater than* x_2*; (2) that it occurs at random between these limits; then the probability that it lies between* a *and* b *is* $(b - a)/(x_2 - x_1)$*, provided, of course, that both* a *and* b *lie within the interval in question.*

The probability that the chosen number is equal to any preassigned number x *is zero.*

In the sections which follow we shall drop our mechanistic ideas as to how the numbers are chosen and shall speak in these more general terms instead.

§ 58. *Distribution Functions for Continuous Variables*

Just as " equally likely events," in spite of their theoretical importance in giving us a starting point for the discussion of discrete groups of events, are comparatively rare in practical studies, so " randomness " (or, if we prefer the phrase, " equally likely intervals ") is also of greater theoretical than practical importance. Most variables which arise in engineering have certain preferred ranges, and other ranges which are exceedingly rare. We have mentioned the resistance of a lamp filament as a quantity which cannot be forecast with absolute certainty; yet if the ideal at which production aims is 300 ohms, it is much more likely that a particular lamp will be

found to lie within the ten-ohm range between 300 and 310 than in the equal range between 390 and 400.

Such a variable is best thought of in connection with its distribution curve—that is, a curve so constructed that the area under it, between the ordinates at $x = a$ and $x = b$, represents the chance of the variable x lying within these limits. Suppose Fig. 12 is such a curve. It follows at once that the total area under the curve, from $-\infty$ to $+\infty$, must be unity, for x is certain to have a value between these limits. Suppose we let this curve have the equation $y = p(x)$, and suppose we denote by $p(> a)$ the chance of x exceeding a,

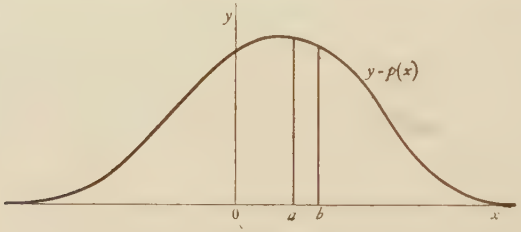


Fig. 12.

and by $p(< a)$ the chance of it being less than a. We see immediately that these three functions are related as follows:

$$p(> a) = \int_a^{\infty} p(x)\, dx,$$

$$p(< a) = \int_{-\infty}^{a} p(x)\, dx.$$

One gives the area to the right of $x = a$, the other the area to the left. In general, the probability of a value of x between a and b is

$$\int_a^b p(x)\, dx.$$

Suppose, now, that b and a are very nearly equal: say, $b = a + da$. If the curve $y = p(x)$ is continuous in the neighborhood of $x = a$, as it usually is in practice, the figure bounded by the x-axis, the curve, and the two ordinates at a

and $a + da$ does not differ much from a rectangle, and hence its area does not differ much from $p(a)\,da$. The difference, in fact, is an infinitesimal of order higher than the first in da, which may usually be ignored. Hence, so long as we deal with very narrow ranges, it is usually quite satisfactory to regard $p(a)\,da$ as the probability of a value occurring within such a range.[1]

As illustrations of distribution functions for variables which are not distributed at random, we give two simple examples. One is a highly artificial case in which the probabilities can be calculated; the other is of the much more common type where they must be inferred as well as may be from the results of a long series of observations.

§ 59. *A Variable which is Not Distributed at Random*

EXAMPLE 38.—*The circumference of the wheel of Example 37 carries a logarithmic scale numbered from 1 to 10. What is the distribution curve for the choice of numbers on this scale?*

Let us denote the *numbers* appearing on the scale by y, and their *distances* from the number 1 by x. Then $x = \log y$. This is the definition of a logarithmic scale.

Now choose two *numbers* a and b. The *distance* between them is $\log b - \log a$; and the circumference of the wheel is $\log 10 - \log 1$. Hence the probability of the pointer indicating a number between a and b is $(\log b - \log a)/(\log 10 - \log 1)$.

If we use logarithms to the base 10, $\log 10 = 1$ and $\log 1 = 0$, so this probability is simply $\log b - \log a$. Our distribution curve, Fig. 13, must therefore be of such a nature that for any interval (b, a) the area

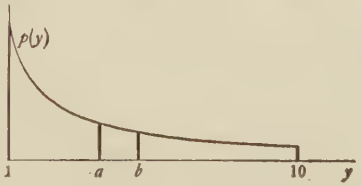


FIG. 13.

[1] It is also not unusual to use the phrase "the probability that x takes the value a is $p(a)$," meaning thereby that the chance of x lying between a and $a + da$ differs from $p(a)\,da$ by an infinitesimal of the second order at least in da; or, in other words, that the ordinate to the distribution curve at a has the length $p(a)$.

is given by this formula. But obviously this means that

$$\int_a^b p(y)\,dy = \log b - \log a.$$

Differentiating this with respect to b we obtain [1]

$$p(b) = \frac{\log e}{b},$$

or, if we prefer to write it that way,

$$p(y) = \frac{\log e}{y},$$

as the definition of our curve.

This case can be solved directly from our definition of probability. Let us now take one upon which we can get no grip theoretically.

§ 60. *Distribution Functions Derived Empirically*

EXAMPLE 39.—*Construct a distribution curve of length of life: in other words a curve the area under any portion of which is a child's chance at birth of dying within the corresponding range of ages.*

It is certainly true that this problem cannot be solved from a purely theoretical standpoint. Equal age ranges are known not to be equally likely, and most of the other facts of mor-

[1] Remember that these are logarithms with base 10. To renew acquaintance with the formula of differentiation in such cases we may point out that, if x is any number whatever,

$$x = e^{\log_e x}.$$

Also

$$e = 10^{\log_{10} e}.$$

Substituting the second in the first

$$x = 10^{(\log_e x)(\log_{10} e)}.$$

But by definition

$$x = 10^{\log_{10} x}.$$

Comparing these, it follows that

$$\log_{10} x = \log_{10} e \log_e x.$$

Differentiating

$$\frac{d}{dx}\log_{10} x = \frac{\log_{10} e}{x}.$$

tality are tempered by experience in the same way. The best we can do is to resort to vital statistics and find what proportion of the population has been observed to die within various age limits. In a sense each life is an "independent trial," and we can conclude from Bernoulli's Theorem that, if the cases are numerous enough, the proportion of deaths between any two age limits is not likely to deviate much from the probability of dying in that range.

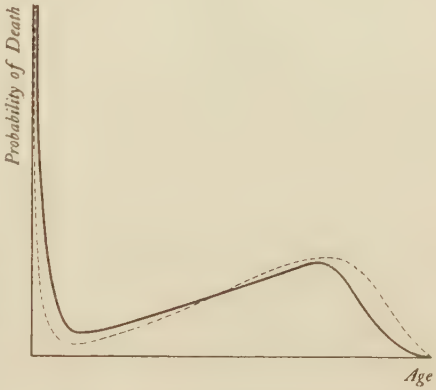


FIG. 14.—A TYPICAL LIFE CURVE.

The curve of Figure 14 has been plotted in this manner from certain German data given by Czuber in his *Wahrscheinlichkeitsrechnung*. The general character of the curve would be much the same for data taken from almost any civilized country; though it is obvious that different conditions of sanitation, and particularly different customs in the handling of children, would affect it somewhat.

The curve is, therefore, of the nature of a "conditional probability" curve: it gives a child's chance of dying at a certain age if he is born at a particular place. It is also conditional in another sense which is of more consequence to the engineer. The conditions of life are not static: medical efficiency, for example, is increasing. Hence the curve depicts a child's chances of life *if born at a given time*. In the very nature of things no child can be born at that time: for the existence of the data implies that the time is past. To-day a child's chances are different, probably in some such way as is indicated by the dotted line, though such a curve can be based on no more substantial foundation than the estimation of a "trend"—or rate at which the probabilities are changing.

The important point to notice, however, is that the data

is unsatisfactory *because* THE PROBABILITIES *change with time.* When this is *not* true, conditions are said to be in "statistical equilibrium."

Present probability can only be inferred from data already collected when the system under consideration is in statistical equilibrium.

Some conception of the significance of the uncertainties introduced into life insurance by this factor can be gotten by considering the position in which the insurance companies would find themselves if the "trend" were toward shorter rather than longer life. They would then face almost certain loss if they based their rates upon available statistics, and would be forced to estimate as best they could the factor by which present conditions differed from past, with obvious evil consequences if they made a wrong guess. As it is they face almost certain gain; which is quite satisfactory to the companies, and not very serious to the policy-holder if his company has his interests at heart.

Now, we have no interest in life insurance as such. But the same conditions are constantly met in engineering. Take, for instance, Example 36, in which certain empirical data were made the foundation for a solution of the problem. Perhaps, in the case in question, such data could be thoroughly relied upon, for screw-making is, I suppose, a pretty stable process, and not subject to rapid improvement. Suppose, however, that the same sort of argument were attempted in the case of some sort of radio accessory: about the time we began to feel some confidence in our data someone would "improve" the process, and we would be confronted with a set of conditions to which our data no longer applied *exactly*—perhaps not at all.

In other words, manufacturing conditions, like life, are not in statistical equilibrum. Instead, *the probabilities are functions of time*, and those functions are generally unknown.

§ 61. *Distribution Functions in Many Variables*

To the conception of " compound events " in the case of discrete variables correspond " functions of several variables " when the variables are continuous. As an example, suppose we are stamping out metal discs on a punch press, to be used in operating a slot machine. If the slot machine is so constructed as to reject a coin the weight of which deviates too much from standard, we will naturally be interested in the probable variation in weight of our product. The weight, however, depends principally upon two factors: the thickness, t, of the particular portion of the sheet from which the disc comes, and its area, a. Both of these are subject to variation, and both are obviously continuous variables.

Suppose we represent a and t as Cartesian coordinates, as in Fig. 15, and build up a " distribution surface"; that is, a surface such that the volume under any portion of it is equal to the probability of the point (a, t) lying under that portion. Call the height of such a surface $p(a, t)$. Then the following things are immediately obvious:

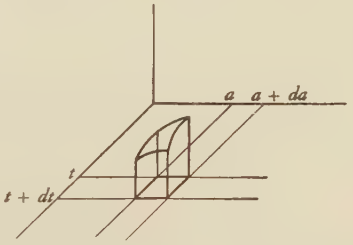


Fig. 15.

1. The probability of a pair of values within the ranges $(a, a + da), (t, t + dt)$ is $p(a,t)\, da\, dt$, except for an infinitesimal of higher order in both da and dt.

2. The probability of a value of a in the range $(a, a + da)$ is equal to the volume of the vertical slab the base of which is bounded by the lines a and $a + da$. If we call it $p(a)\, da$, it is related to $p(a, t)$ by the law [1]

$$p(a)\, da = da \int p(a, t)\, dt.$$

[1] " Except for a differential of higher order in da." The general idea should by now be clear enough to allow the omission of such statements in the future. Differential notation implies a limiting process for complete accuracy.

3. The probability of a value of t between t and $t + dt$ is represented by a similar slab. The corresponding formula is

$$p(t) = \int p(a, t)\, da. \tag{56}$$

It is understood, of course, that the integral extends over every possible value of a.

4. We can, if we wish, think of the occurrence of a value of a in the slab between a and $a + da$ as "event A," and the occurrence of a value of t between t and $t + dt$ as "event B." To these events the argument of § 47 applies directly, wherefore we have

$$P(AB) = P(A)\, P_A(B). \tag{20}$$

But we already know that $P(A)$ differs from $p(a)\, da$ by an infinitesimal of higher order in da, which we may formulate in the form of an equation by writing

$$P(A) = [p(a) + \epsilon]\, da,$$

it being understood that ϵ vanishes with da. Similarly we know that $P(AB)$ differs from $p(a, t)\, da\, dt$ by an infinitesimal of higher order than $da\, dt$. Hence

$$P(AB) = [p(a, t) + \epsilon']\, da\, dt,$$

where ϵ' also vanishes.

Introducing these expressions into (20) and making a few obvious rearrangements we obtain

$$\frac{P_A(B)}{dt} = \frac{p(a, t) + \epsilon'}{p(a) + \epsilon}.$$

Now let da and dt approach zero. The right-hand side of the equation approaches the limit $p(a, t)/p(a)$, which we may call $p_a(t)$ if we like. Therefore the left-hand side must also approach this same limit. This means, of course, that $p_a(t)\, dt$ differs from $P_A(B)$ by an infinitesimal of higher order than dt. We are therefore justified in calling $p_a(t)\, dt$ "the conditional probability of the point lying between t and

$t + dt$, if it is known to lie between a and $a + da$"; meaning, as in all such differential expressions, that a differential of higher order is ignored.

However, the very definition of $p_a(t)$ shows that it satisfies the relation

$$p(a, t) = p(a)\, p_a(t), \tag{57}$$

which is the exact counterpart of (20) in the case of continuous variables.

If we substitute (57) in (56) we get

$$p(t) = \int p(a)\, p_a(t)\, da, \tag{58}$$

which is an exact counterpart of (21).

5. Finally, if we had started from the slab for which t lies between t and $t + dt$ we would have concluded that

$$p(a, t) = p(t)\, p_t(a);$$

and thereafter, by comparing this equation with (57) and making certain obvious transformations, that

$$p_t(a) = \frac{p(a)\, p_a(t)}{p(t)}, \tag{59}$$

or from (58)

$$p_t(a) = \frac{p(a)\, p_a(t)}{\int p(a)\, p_a(t)\, da}. \tag{60}$$

These are extensions of (53) and (54). That is, they are Bayes' Theorem for the case of continuous variables.

Obviously, the significance of "thickness" and "area" which we have attached to t and a is unimportant: the formulæ are true in general. Nor are the ideas confined to the case of two variables only. Entirely similar formulæ could be written for any number of variables. For instance, the velocity of a gas molecule may be defined in terms of its components in three coordinate directions x, y, z. There is, of course, a

certain probability of a molecule possessing this velocity.[1] We denote it by $p(u, v, w)$; u, v, w being the three components in question.

Of course we mean by this, that the chance that all three components lie within the ranges $(u, u + du)$, $(v, v + dv)$, $(w, w + dw)$ is $p(u, v, w)\, du\, dv\, dw$. The geometric picture, however, is no longer quite so simple as before, for it requires *three* dimensions to depict the variables u, v, w, and a *fourth* would be needed for p. As we cannot visualize four dimensional space we can only proceed by analogy; and in doing this it is simpler to speak of "integrating p throughout a volume" than to try to visualize the fourth dimension. Thus, referring again to the case of two variables, (56) tells us that $p(a)\, da$ is the result of "integrating $p(a, t)$ over the vertical strip bounded by the lines a and $a + da$." In this language we say that the probability of the point (u, v, w) lying within any closed volume is obtained by integrating $p(u, v, w)$ over the volume in question. It is not necessary that this volume be bounded by coordinate planes. It is entirely unrestricted.

§ 62. *Some Examples of Change of Variable in Distribution Functions*

In studying gas molecules we are not always interested in their velocity. We may want to discuss speed, direction, momentum, energy—any number of things. But as all these are determined by the velocity,[2] we ought to be able to find the probability of any of them in terms of $p(u, v, w)$. This is really the case.

For example, suppose we are required to find the chance

---

[1] We are using the term "velocity" in its scientific sense of "speed and direction," not merely "speed." Two bodies, one moving along the x-axis at the rate of one centimeter per second, the other at the same rate along the y-axis, have the same "speed" but different "velocities." To say that two bodies have the same "velocity" implies that all three components are equal.

In other words, velocity is a vector.

[2] And the mass, which is generally a constant.

that a molecule has a speed less than s. This means that u, v, w must have such values that $u^2 + v^2 + w^2 < s^2$; in other words, that the point (u, v, w) must lie inside the sphere of radius s about the origin. Thus

$$p(< s) = \iiint p(u, v, w)\, du\, dv\, dw,$$

the limits of integration being the surface of the sphere.

Or, in the two-dimensional case, suppose we want the probability that the weight w lies between w_0 and $w_0 + dw$. We have already noted that[1] $w = ta$. Hence t and a must lie inside the strip bounded by

$$ta = w_0,$$

$$ta = w_0 + dw.$$

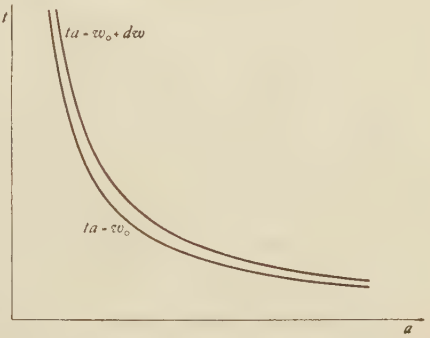


Fig. 16.

These are the hyperbolas shown in Fig. 16. Hence

$$p(w)\, dw = \int_0^\infty da \int_{\frac{w_0}{a}}^{\frac{w_0+dw}{a}} p(a, t)\, dt.$$

We are supposing that dw is very small. If so, the limits of integration upon t are very close together, except when a itself is exceedingly small. An exceedingly small a, however, means (in our problem) a disc of negligible area. Obviously such discs are so unlikely that we shall commit no practical error in asserting that they never occur. Let us say that discs of area less than a_0 never occur. Then the t-limits are *always* very close together, if dw is small.

Further, if $p(a, t)$ is continuous, its value will not differ

[1] Usually there would be a density factor. We may suppose the unit of weight to be so chosen, however, that this factor is unity.

much from $p(a, w_0/a)$ over the entire range of t-integration. Hence

$$\int_{\frac{w_0}{a}}^{\frac{w_0+dw}{a}} p(a,t)\,dt \doteq p\left(a, \frac{w_0}{a}\right)\frac{dw}{a}$$

and

$$p(w_0)\,dw \doteq dw\int_0^\infty p\left(a, \frac{w_0}{a}\right)\frac{da}{a}.$$

The approximation is within the second order of the infinitesimal dw, and since our interpretation of $p(w_0)$ implies either a limiting process or an error of that magnitude, it is allowable to replace $\doteq$ by $=$. Hence, finally, dropping the subscript zero which really means nothing to us,

$$p(w) = \int_0^\infty p\left(a, \frac{w}{a}\right)\frac{da}{a}. \tag{61}$$

We shall wish to refer to this result again, and it will prove wise to call attention at this time to a false argument which we might thoughtlessly have carried out. It is this: to demand that w take the value w_0 is equivalent to demanding that t take the value w_0/a. But by (56) this is

$$\int p\left(a, \frac{w_0}{a}\right)da.$$

This, however, is *not* identical with the correct answer (61). The reader will have no difficulty in locating the error.

Both these illustrations may be classified as "change of variable": we have shifted our attention from u, v, w to s; or from t, a to w. In each case we have at the same time reduced the number of variables, though we need not have done so. We might, if we had desired, have asked for the probability of a weight w and a radius r: two new variables rather than one. Such changes of variable are obviously to be expected as our attention shifts from one phase of a problem to another; and as we have seen that a simple substitution of

new variables for old in the equation of p does not give correct results, we must investigate the matter in detail.

§ 63. *Change of Variable in Distribution Functions*

Let us consider first a case in which only one variable is involved, and think in terms of Examples 37 and 38. In solving 38 we really made use of our knowledge of the solution of 37: that is, we knew $p(x)$ and the relation $x = \log y$, and from it found $p(y)$.

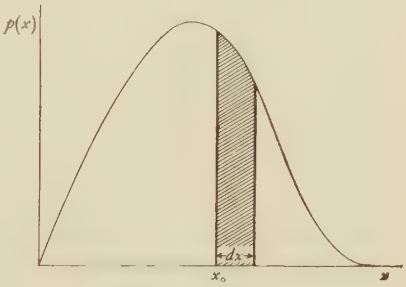


Fig. 17.

Let us suppose that $p(x)$ is represented by the curve of Fig. 17. In Example 38, of course, it was a straight line parallel to the x-axis; but as we wish to make sure that our final result is perfectly general, we choose a better picture of "any function." Suppose also that $p(y)$ has somehow been found, and is represented by the curve of Fig. 18. Finally, suppose the relation between y and x is known in the form of an equation

$$y = f(x), \qquad (62)$$

just as in Example 38 it was known in the form $y = 10^x$.

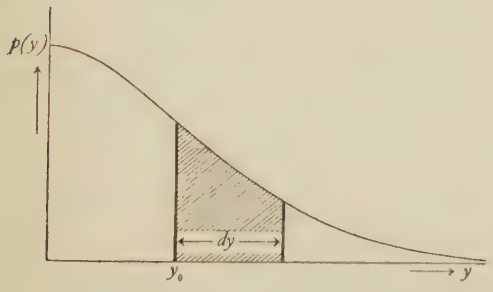


Fig. 18.

The chance that x lies between two values x_0 and $x_0 + dx$ is represented by the shaded area in Fig. 17. But whenever x lies within this interval, y also must lie within fixed limits. It will, in fact, lie between $y_0 = f(x_0)$ and $y_0 + dy = f(x_0 + dx)$; that is, within the shaded area of

Fig. 18. As neither event can happen unless the other does, these shaded areas must be equal.

Now suppose that dx is very small. The same will also be true of dy.[1] Hence both shaded areas are nearly rectangular. If, then, h and b denote height and breadth,

$$h_x b_x = h_y b_y. \tag{63}$$

But the heights are by definition $p(x)$ and $p(y)$, while the breadths are dx and $f(x_0 + dx) - f(x_0)$, the latter of which is approximately[2] $f'(x_0)\,dx$. Hence, dropping the subscripts, which are no longer of value to us, we find the equation of the new distribution curve to be

$$p(y) = \frac{p(x)}{f'(x)}. \tag{64}$$

Of course the y and x in these equations are *corresponding* values. In Example 38, for instance, the y is always 10^x and conversely x is always $\log y$. Hence for this special problem (64) becomes

$$p(y) = \left.\frac{p(x)}{\frac{d}{dx}(10^x)}\right|_{x = \log y},$$

which reduces to the answer obtained in § 59 when we recall that $p(x) = 1$ and that the derivative of 10^x is $10^x/\log e$.

Similarly, in (64) the proper x may be obtained by solving (62). As it depends on y for its value, we may properly call it[3] $x = f^{-1}(y)$, and write (64) as

$$p(y) = \left.\frac{p(x)}{f'(x)}\right|_{x = f^{-1}(y)}. \tag{65}$$

This, then, is a general formula by means of which, when the distribution function for any variable x is known, together with

[1] Unless dy/dx, as determined from (62), is infinite at $x = x_0$.

[2] Under the conditions of the last note.

[3] The expression for x in terms of y is called the "inverse" of the expression for y in terms of x, and is always denoted by the index -1. Thus the inverse of the $y = f(x) = 10^x$ in § 59 is $x = f^{-1}(y) = \log y$.

the relationship between x and some other variable y, the distribution function for y may be obtained. It says nothing whatever, except that the curves must be so defined that corresponding elements of area are equal.

Similar rules may be set up for distribution functions of several variables. To illustrate how this is done, let us consider first the case of two variables x and y. To represent these requires a plane; any pair of values of x and y determine a point (x, y) somewhere in the plane, and the distribution function $p(x, y)$ is defined as a function which, when integrated over a region dA, gives the probability that the point (x, y) lies within dA. If dA is small enough, $p(x, y)$ is sensibly constant over the entire area, and the probability is given by [1] $p(x, y)\, dA$.

Now suppose we have two other variables ξ and η, related to x and y by means of the equations

$$\left.\begin{aligned} \xi &= f(x, y), \\ \eta &= \phi(x, y), \end{aligned}\right\} \tag{66}$$

corresponding to the equation $y = f(x)$ in the case of one variable. Suppose we choose another plane for the representation of these, and seek to determine their distribution function $p(\xi, \eta)$. As the point (x, y) travels around the boundary of dA, the corresponding point (ξ, η) will travel around some curve in its own plane. This curve will bound an area $d\alpha$, which we shall call the " element of area corresponding to dA." Whenever the point (x, y) lies in dA, (ξ, η) lies in $d\alpha$. Therefore

$$p(\xi, \eta)\, d\alpha = p(x, y)\, dA. \tag{67}$$

If x and y are given in advance, the ξ and η of this equation are to be found from (66). Conversely, if ξ and η are given,

[1] In particular, if dA is a rectangle bounded by horizontal and vertical lines its area may be called $dx\, dy$. The probability in question will then be $p(x, y)\, dx\, dy$. But as it is by no means necessary that the area be of such a shape, we prefer to keep the more general differential dA.

the x and y must be found in terms of ξ and η by solving (66) just as, in writing (65), $y = f(x)$ had to be solved for x.

As has already been said dA and $d\alpha$ are *corresponding* elements of area. To get $p(\xi, \eta)$, therefore, we must multiply $p(x, y)$ by the quotient $dA/d\alpha$ just as in (64) we multiplied $p(x)$ by the quotient $dx/dy = 1/f'(x)$.

To find this quotient we choose dA as a rectangle bounded by the values x, $x + dx$, y and $y + dy$. (See Fig. 19.) As (x, y) travels along one of the lines which bound this rectangle (ξ, η) will trace a curve of some form. The element $d\alpha$, therefore, is bounded by four curved elements as shown in the

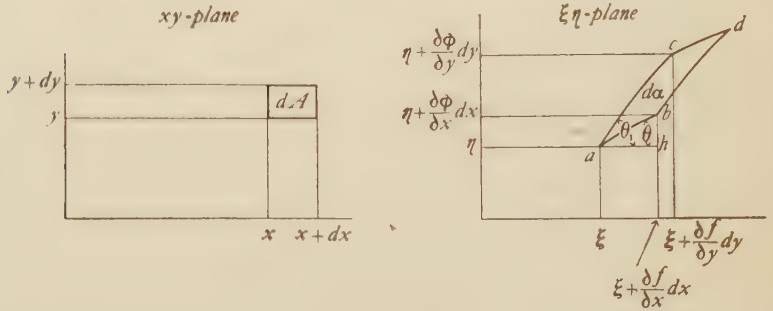


Fig. 19.

figure. If dx and dy are very small these elements will be so nearly straight that the difference between the true $d\alpha$ and the area of the rectilinear figure having the corners a, b, c, d can be ignored. Hence it is sufficient for the purpose of our argument to locate the four points in the $\xi\eta$-plane which correspond to the corners of dA. We may then assume that they are connected by straight lines.

We consider first the point corresponding to (x, y); that is, to the lower left-hand corner of dA. Let us suppose that this is the point a of Fig. 19.[1] Its coordinates are given by (66).

Next we take $(x + dx, y)$ — the lower right-hand corner of dA. The corresponding corner b of $d\alpha$ has the coordinates

[1] It might be any corner. It all depends on the signs of the derivatives of f and ϕ.

$f(x + dx, y)$ and $\phi(x + dx, y)$. If dx is small enough these may be replaced by [1]

$$f(x, y) + \frac{\partial f}{\partial x} dx \quad \text{and} \quad \phi(x, y) + \frac{\partial \phi}{\partial x} dx,$$

respectively. But $f(x, y)$ is ξ, and $\phi(x, y)$ is η; that is, these are the coordinates of the point a. Hence $\frac{\partial f}{\partial x} dx$ and $\frac{\partial \phi}{\partial x} dx$ are the horizontal and vertical sides of the triangle ahb.

Next we consider the upper left-hand corner $(x, y + dy)$. In the $\xi\eta$-plane this becomes a point of which the coordinates are easily found to be $\xi + \frac{\partial f}{\partial y} dy$ and $\eta + \frac{\partial \phi}{\partial y} dy$. It is represented by c in Fig. 19.

We do not need to know the coordinates of the fourth point d, for if dx and dy are very small, $d\alpha$ is very nearly a parallelogram, and therefore its area is satisfactorily determined by the two sides ab and ac. The area is, in fact, equal to the product of the lengths ab and ac and the sine of the angle included between them. In terms of the angles θ and θ_1 this rule gives

$$d\alpha = ab \cdot ac \sin(\theta_1 - \theta)$$

$$= (ab \cos\theta)(ac \sin\theta_1) - (ab \sin\theta)(ac \cos\theta_1).$$

However

$$ah = ab \cos\theta = \frac{\partial f}{\partial x} dx,$$

$$bh = ab \sin\theta = \frac{\partial \phi}{\partial x} dx,$$

$$ac \cos\theta_1 = \frac{\partial f}{\partial y} dy,$$

$$ac \sin\theta_1 = \frac{\partial \phi}{\partial y} dy;$$

[1] Partial differential notation is used because y remains constant along this side of our rectangle.

wherefore

$$d\alpha = \left(\frac{\partial f}{\partial x}\frac{\partial \phi}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial \phi}{\partial x}\right) dA.$$

Thus we have found the ratio of $d\alpha$ to dA. It serves our purpose best to write it in the form of a determinant:

$$d\alpha = \begin{vmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial \phi}{\partial x} & \dfrac{\partial \phi}{\partial y} \end{vmatrix} dA.$$

As f and ϕ really mean the same thing as ξ and η [we could, in fact, have written (66) in the form $\xi = \xi(x, y)$, $\eta = \eta(x, y)$, if we had wished], this determinant can just as well be written

$$\frac{\partial(\xi, \eta)}{\partial(x, y)} = \begin{vmatrix} \dfrac{\partial \xi}{\partial x} & \dfrac{\partial \xi}{\partial y} \\ \dfrac{\partial \eta}{\partial x} & \dfrac{\partial \eta}{\partial y} \end{vmatrix}. \tag{68}$$

Then, by (67),

$$p(\xi, \eta) = \frac{p(x, y)}{\dfrac{\partial(\xi, \eta)}{\partial(x, y)}},$$

it being understood, of course, that the x's and y's on the right-hand side are to be expressed in terms of ξ and η by solving (66).

If we were dealing with a problem in which there were three variables, x, y and z, we should be forced to talk about elements of volume instead of elements of area; and by an exactly similar line of argument we could determine the ratio of the volumes of corresponding elements in $\xi\eta\zeta$-space and xyz-space. We shall not go through the argument involved in this case as it is identical in principle with that used above. The result, however, is

$$d\alpha = \begin{vmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} & \frac{\partial \xi}{\partial z} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} & \frac{\partial \eta}{\partial z} \\ \frac{\partial \zeta}{\partial x} & \frac{\partial \zeta}{\partial y} & \frac{\partial \zeta}{\partial z} \end{vmatrix} dA,$$

the letters A and α being retained for corresponding elements in the xyz- and $\xi\eta\zeta$-spaces.

In general, no matter how many variables $x, y, z, \ldots, w$ may be involved, when we change from one system of representation to another the new differential element $d\alpha$ is defined in terms of the old element dA by the equation

$$d\alpha = \begin{vmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} & \frac{\partial \xi}{\partial z} & \cdots & \frac{\partial \xi}{\partial w} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} & \frac{\partial \eta}{\partial z} & \cdots & \frac{\partial \eta}{\partial w} \\ \frac{\partial \zeta}{\partial x} & \frac{\partial \zeta}{\partial y} & \frac{\partial \zeta}{\partial z} & \cdots & \frac{\partial \zeta}{\partial w} \\ \cdots & \cdots & \cdots & & \cdots \\ \frac{\partial \omega}{\partial x} & \frac{\partial \omega}{\partial y} & \frac{\partial \omega}{\partial z} & \cdots & \frac{\partial \omega}{\partial w} \end{vmatrix} dA, \qquad (69)$$

or, symbolically,

$$d\alpha = \frac{\partial(\xi, \eta, \zeta, \ldots, \omega)}{\partial(x, y, z, \ldots, w)} dA.$$

Hence

$$p(\xi, \eta, \zeta, \ldots, \omega) = \frac{p(x, y, z, \ldots, w)}{\dfrac{\partial(\xi, \eta, \zeta, \ldots, \omega)}{\partial(x, y, z, \ldots, w)}}. \qquad (70)$$

The determinant $\dfrac{\partial(\xi, \eta, \zeta, \ldots, \omega)}{\partial(x, y, z, \ldots, w)}$, in which the elements are all possible partial derivatives of the new variables with respect to the

old, is known as the "Jacobian" of the transformation. Its geometrical significance is contained in the concept which has led us to it: the ratio of the differential elements in the two systems of coordinates. About it we may make three interesting observations:

In the first place, it is interesting to note that the simple case discussed at the beginning of the section, in which we had only one new variable y and one old variable x, falls in line with this more elaborate equation. As only *one relation* $y = f(x)$ was required to define the new variable, only one *row* appears in the determinant (69); and since there is only *one variable* x in the original system, there can be only one derivative of f, and therefore only one *column* in (69). The entire Jacobian therefore reduces to the single term df/dx, which is therefore the ratio of our old and new differential elements. This is, in fact, what we found it to be.

In the second place, there is a perfect reciprocal relationship between our old and new variables. Perhaps this is most easily realized by recalling that, once $p(\xi, \eta, \zeta, \ldots, \omega)$ has been found, it is a "known distribution function," and we could start with it and find $p(x, y, z, \ldots, w)$ by exactly the same argument as was used above. The only essential difference would be, that instead of differentiating the equations $\xi = \xi(x, y, z, \ldots, w)$, $\eta = \eta(x, y, z, \ldots, w), \ldots,$ to form the Jacobian, we should want equations defining $x, y, z, \ldots, w$ in terms of $\xi, \eta, \zeta, \ldots, \omega$. Obviously the result of this reciprocal relationship is

$$p(x, y, z, \ldots, w) = \frac{p(\xi, \eta, \zeta, \ldots, \omega)}{\dfrac{\partial(\xi, \eta, \zeta, \ldots, \omega)}{\partial(x, y, z, \ldots, w)}}. \tag{71}$$

Now the p's mean exactly the same things in (70) and (71). Hence we obtain the following remarkable (though, considering its geometrical significance, entirely reasonable) property of Jacobians: that

$$\frac{\partial(x, y, z, \ldots, w)}{\partial(\xi, \eta, \zeta, \ldots, \omega)} = \begin{vmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial x}{\partial \eta} & \cdots & \dfrac{\partial x}{\partial \omega} \\ \dfrac{\partial y}{\partial \xi} & \dfrac{\partial y}{\partial \eta} & \cdots & \dfrac{\partial y}{\partial \omega} \\ \cdots & \cdots & & \cdots \\ \dfrac{\partial w}{\partial \xi} & \dfrac{\partial w}{\partial \eta} & \cdots & \dfrac{\partial w}{\partial \omega} \end{vmatrix}$$

is the reciprocal of

$$\frac{\partial(\xi, \eta, \zeta, \ldots, \omega)}{\partial(x, y, z, \ldots, w)} = \begin{vmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} & \cdots & \frac{\partial \xi}{\partial w} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} & \cdots & \frac{\partial \eta}{\partial w} \\ \cdots & \cdots & & \cdots \\ \frac{\partial \omega}{\partial x} & \frac{\partial \omega}{\partial y} & \cdots & \frac{\partial \omega}{\partial w} \end{vmatrix}.$$

Suppose, for instance, that x and y are related to ξ and η by the equations

$$x = \xi \cos \eta,$$
$$y = \xi \sin \eta.$$

Then

$$\frac{\partial(x, y)}{\partial(\xi, \eta)} = \begin{vmatrix} \cos \eta & -\xi \sin \eta \\ \sin \eta & \xi \cos \eta \end{vmatrix} = \xi.$$

But by solving for ξ and η we find

$$\xi = \sqrt{x^2 + y^2},$$
$$\eta = \tan^{-1} \frac{y}{x};$$

whence

$$\frac{\partial(\xi, \eta)}{\partial(x, y)} = \begin{vmatrix} \frac{x}{\sqrt{x^2 + y^2}} & \frac{y}{\sqrt{x^2 + y^2}} \\ \frac{-y}{x^2 + y^2} & \frac{x}{x^2 + y^2} \end{vmatrix} = \frac{1}{\sqrt{x^2 + y^2}} = \frac{1}{\xi} = \frac{1}{\frac{\partial(x, y)}{\partial(\xi, \eta)}}.$$

Our third observation is, that, since the ratio of two volumes cannot be negative,[1] the determinant which occurs in (69) should always be positive. Whether it is or not, however, depends upon the order in which the variables are written down; for we can always change the sign of a determinant by interchanging two rows or two

[1] In other branches of mathematics, particularly in Analysis Situs, the *sign* of the Jacobian has an important significance. As we are obviously dealing with absolute magnitudes, however, it is unnecessary to discuss the matter here.

columns. If, for example, we had happened to write the equations of the preceding paragraph in the order

$$y = \xi \sin \eta,$$
$$x = \xi \cos \eta,$$

we would have found that $\dfrac{\partial(y, x)}{\partial(\xi, \eta)} = -\xi$.

It follows, therefore, that the sign with which our Jacobian appears is largely a matter of accident, and that in every case the positive sign is the correct one.

As a further illustration of this method, let us reconsider the problem solved in § 62. In that problem we supposed the distribution function to be known for the variables a and t, and desired a new distribution function for the single variable w, which was related to a and t by the equation

$$w = ta.$$

Let us first attempt to find a new distribution function in terms of the *two* new variables

$$w = ta,$$
$$v = a.$$

Then by actual evaluation of the Jacobian, we find that

$$\frac{\partial(w, v)}{\partial(a, t)} = \begin{vmatrix} t & a \\ 1 & 0 \end{vmatrix} = -a.$$

Hence, disregarding the negative sign, we arrive at the new distribution function [1]

$$P(w, v) = \left. p(a, t) \right|_{\substack{a = v \\ t = w/v}} = \frac{p\left(v, \dfrac{w}{v}\right)}{v}.$$

Finally from (56) we get a distribution function for the single variable w in the form

$$p(w) = \int_0^\infty P(w, v)\, dv$$
$$= \int_0^\infty p\left(v, \frac{w}{v}\right) \frac{dv}{v};$$

[1] The capital P is used to avoid confusion with the last member of the equation.

which is, except for the introduction of a new symbol for the variable of integration, the same as (61).

§ 64. *Derivation of a Distribution Function for the Velocities of Gas Molecules*

One of the important problems in Physics to which Probability Theory was first applied, and virtually the only one in which the so-called Normal Law appears as a consequence of an argument which is even approximately sound physically, is that of the distribution of velocities within a perfect gas. Later on, in § 147, we shall give what is probably its most satisfactory proof. For the present we wish to use it principally as an illustration of the processes which we have been discussing in the present chapter, and shall therefore content ourselves with a "proof" that is far from exact.

We suppose the gas to be composed of a large number of particles, similar or dissimilar as the case may be, but all in agitated motion. We further assume that the gas is restrained within a vessel which is not itself in motion. The question before us is, What is the chance of a particular particle having a specified velocity?

To start with, we want to be sure that we know exactly what we mean by this question. Suppose we phrase it this way: A particular particle is tagged for purposes of identification, and an instant of observation is chosen without any advance information which influences the probability for which we seek. At that instant we note the velocity of the particle in question, both as to magnitude and direction. What is the chance that its components lie between u and $u + du$, v and $v + dv$, and w and $w + dw$, respectively?

We make these assumptions:

(*a*) The probability of a given velocity is independent of the part of the vessel in which the particle may be. That is, $p^*(u, v, w)$ does not depend upon the coordinates x, y, z.[1]

[1] The asterisk is added to the symbol for probability in order to conform to the notation of Chapter XI.

Physically there are two common conditions which violate this assumption: when different parts of the gas are at different temperatures; and when they are at different pressures as, for example, when it is flowing out of an orifice.

(*b*) All directions of motion are equally likely.

To this we cannot object.

(*c*) The probability of a component u in the x-direction is independent of any knowledge we may have about the transverse components v and w.

In this assumption lies the weakness of the argument; for it is certainly not obvious that it should be true, either from a mathematical or a physical point of view. Our "proof" therefore will not be very convincing. However, we proceed with it.

Let $p^*(u)$ be the probability of a component u. By the assumption (*c*):

$$p^*(u, v, w) = p^*(u) \cdot p^*(v) \cdot p^*(w),$$

or

$$\log p^*(u, v, w) = \log p^*(u) + \log p^*(v) + \log p^*(w).$$

By assumption (*b*) the function $p^*(u, v, w)$ must depend on the combination $u^2 + v^2 + w^2$ only; for if we were to replace the three components of velocity by their expressions in terms of speed and direction, the directional variables would all disappear, leaving only the speed. Suppose, then, that we write for the moment

$$\log p^*(u, v, w) = f(u^2 + v^2 + w^2).$$

We now make our final assumption: that this function f is of such a nature that it can be expanded in a series for sufficiently small values of the speed. We then get

$$\begin{aligned} f(u, v, w) &= a_0 + a_1(u^2 + v^2 + w^2) + a_2(u^2 + v^2 + w^2)^2 + \ldots \\ &= \log p^*(u) + \log p^*(v) + \log p^*(w). \end{aligned} \tag{72}$$

Obviously the third member of this equation is of such a form that no cross-products between u, v, w can be allowed. But

such cross-products arise from all the terms of the second member except $a_0 + a_1(u^2 + v^2 + w^2)$: hence the coefficients $a_2, a_3, \ldots$, must be zero. Thus

$$\log p^*(u, v, w) = a_0 + a_1(u^2 + v^2 + w^2),$$

or

$$p^*(u, v, w) = A\, e^{a_1(u^2+v^2+w^2)}.$$

This equation contains two constants A and a_1, but they are not both arbitrary because of the necessary relation

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p^*(u, v, w)\, du\, dv\, dw = 1,$$

which requires that $\pi^3 A^2 = -\, a_1{}^3$. If, then, we call $a_1 = -\, a$ we have as our final law [1]

$$p^*(u, v, w) = \left(\frac{a}{\pi}\right)^{3/2} e^{-a(u^2+v^2+w^2)}. \tag{73}$$

This is Maxwell's Equation. It contains only *one* arbitrary constant a, and that can be shown to be completely determined by the temperature of the gas and the mass of the particle under consideration.

§ 65. *Some Instructive Illustrations; Change of Variable in Maxwell's Equation*

As an illustration of the type of service which change of variable may render us, let us consider the following example:

EXAMPLE 40.—*The probability of a gas molecule having the velocity components u, v, w being given by (73), find the distribution law in terms of the variables s (speed), ϕ (co-latitude) and θ (longitude).*

---

[1] It is interesting historically to note that Maxwell first derived equation (73) by substantially this line of argument. He called attention to the fact, however, that there was no real justification for it, and afterwards attempted to improve his demonstration by a line of argument similar to the one given in § 144. I believe he was of the opinion that this latter proof was conclusive, and that it was some years before attention was called to the fact that it also implies an assumption for which there is little greater justification than there is for the assumption that the velocity components are independent.

The coordinate system specified in this example is no other than the one usually known as "spherical coordinates." It is shown in Fig. 20. The relations between the two sets of variables are immediately seen to be

$$\left.\begin{aligned} u &= s \sin\phi \cos\theta, \\ v &= s \sin\phi \sin\theta, \\ w &= s \cos\phi. \end{aligned}\right\} \qquad (74)$$

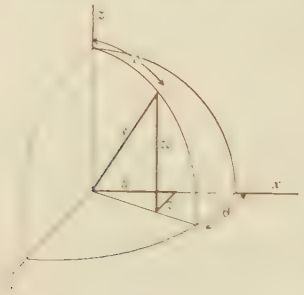


FIG. 20.—SPHERICAL COORDINATES.

Knowing these equations we are prepared at once to transform (73) into our new variables s, ϕ and θ. It happens, of course, that these equations give the *known* variables in terms of the *desired* ones; but that is not a disadvantage, for as we have seen, the Jacobian of one set is the reciprocal of the Jacobian of the other.

We obtain directly

$$\frac{\partial(u, v, w)}{\partial(s, \phi, \theta)} = \begin{vmatrix} \sin\phi\cos\theta & s\cos\phi\cos\theta & -s\sin\phi\sin\theta \\ \sin\phi\sin\theta & s\cos\phi\sin\theta & s\sin\phi\cos\theta \\ \cos\phi & -s\sin\phi & 0 \end{vmatrix}$$

$$= s^2 \sin\phi;$$

and therefore

$$p^*(s, \phi, \theta) = s^2 \sin\phi \, p(u, v, w),$$

or, since $u^2 + v^2 + w^2 = s^2$,

$$p^*(s, \phi, \theta) = \left(\frac{a}{\pi}\right)^{3/2} s^2 \sin\phi \, e^{-as^2}. \qquad (75)$$

This equation is exactly equivalent to (73), but it is expressed in terms of the polar representation of velocities instead of the Cartesian representation.

This equation reveals a treacherous point in our argument. By assumption (*i*), § 64, all directions should be equally likely; and we have already remarked that because of that assumption we were

justified in assuming $p^*(u, v, w)$ to be a function of s^2 only. Yet (75) contains ϕ as well as s^2, and on first thought would seem to violate one of the assumptions upon which it is based.

The explanation is a simple one. By the brief expression "all directions are equally likely" we really meant this sensible thing: that if a direction were named, the chance of our particle deviating from that direction by an angle less than $d\alpha$, say, would not depend upon the chosen direction. Or, in still other words, if at some instant lines were drawn through the particle in any two directions, and if like cones were instantaneously described about both, the chance of the direction of motion of the particle being within one cone would equal the chance of it being in the other.

Now, if a sphere were described about the common apex of these two cones, each cone would cut out an area on the sphere, and obviously these two areas would be equal. Hence assumption (*b*) requires the probability of the direction of motion intersecting this sphere within a certain area to be the same no matter where on the surface of the sphere the area may occur. The element $d\phi\, d\theta$, however, is not of the same area at all places: indeed the element of area is just $r^2 \sin\phi\, d\phi\, d\theta$, if r is the radius; and if we define "element of direction" as the solid angle subtended by this area, the "element of direction" is just $\sin\phi\, d\phi\, d\theta$. Hence the form of (75), which may be written

$$p^*(s, \phi, \theta)\, ds\, d\phi\, d\theta = \left(\frac{a}{\pi}\right)^{3/2} s^2\, e^{-as^2}\, ds\, (\sin\phi\, d\phi\, d\theta),$$

is really in accordance with assumption (*b*).

§ 66. *Information that Can be Derived from (73) and (75)*

From (75) we can easily find the probability that the *speed* of a molecule lies between s and $s + ds$, if we do not care what direction is associated with that speed. We need only integrate $p^*(s, \phi, \theta)\, ds\, d\phi\, d\theta$ over all possible values of θ and ϕ — that is, we need only sum up, for every possible direction, the probability that the desired speed s occurs *in combination with* that direction. However, to include every possible direction we must allow θ to vary from 0 to 2π, and ϕ from 0 to π. Hence

$$p^*(s)\, ds = \left(\frac{a}{\pi}\right)^{3/2} s^2\, e^{-as^2}\, ds \int_0^{\pi} d\phi \int_0^{2\pi} d\theta \sin\phi,$$

Integration gives at once

$$p^*(s) = 4\sqrt{\frac{a^3}{\pi}}\, s^2\, e^{-as^2}. \tag{76}$$

The essential thing to remember about these formulæ is, that (73) and (75) give the probability of a molecule having a specified *velocity* (direction included), while (76) gives the probability of it having a definite *speed.*

Returning again to (73) we ask for the probability that a molecule has u as its component of velocity in the x-direction, regardless of the transverse velocity components with which u may be associated. We obtain this answer from (73) by exactly the same process as was used in obtaining (76) from (75). That is, we sum up $p^*(u, v, w)\ du\ dv\ dw$ for every value of v and w which can possibly be associated with the u in question, thereby obtaining

$$p^*(u) = \int_{-\infty}^{\infty} dv \int_{-\infty}^{\infty} dw\, p^*(u, v, w)$$
$$= \left(\frac{a}{\pi}\right)^{3/2} e^{-au^2} \int_{-\infty}^{\infty} e^{-av^2} dv \int_{-\infty}^{\infty} e^{-aw^2} dw.$$

Each of the integrals involved in this expression is in exactly the same form. If we evaluate one of them, therefore, we shall automatically have the value of the other also. We choose to deal with the one in w. Since e^{-aw^2} is an even function, its integral between the limits $-\infty$ and 0 must be equal to its integral from 0 to $+\infty$. We may therefore compute the integral over the latter range and double its value. However, by means of the substitution $y = aw^2$ we obtain

$$2\int_0^{\infty} e^{-aw^2} dw = \frac{1}{\sqrt{a}} \int_0^{\infty} y^{-1/2} e^{-y} dy;$$

and this integral in y is by definition $-\frac{1}{2}!$, the value of which is known to be $\sqrt{\pi}$. Hence each of the integrals which we

have to evaluate is $\sqrt{\pi/a}$, and their product is π/a. This leads to

$$p^*(u) = \sqrt{\frac{a}{\pi}}\, e^{-au^2}. \tag{77}$$

Since (73) is symmetrical in u, v and w we may write at once the distribution functions for the v and w components of velocity,

$$p^*(v) = \sqrt{\frac{a}{\pi}}\, e^{-av^2},$$

$$p^*(w) = \sqrt{\frac{a}{\pi}}\, e^{-aw^2}.$$

Were we not interested in this place in illustrating the formal operations by means of which distribution functions can be manipulated from one form to another, we could have obtained (77) much more easily from (72). It is at once obvious that $\log p^*(u)$ must be the sum of $a_1 u^2$ and some constant. Moreover $\log p^*(v)$ and $\log p^*(w)$ must also contain this same constant, which therefore is $a_0/3$. This leads at once to (77).

Another type of question which arises frequently in physical problems concerns the sort of molecules which pass through a given surface. For instance, we may be thinking of diffusion of a gas through a hole in the wall of the containing vessel, or we may be thinking of the molecules which pass across an imaginary mathematical surface somewhere inside the vessel. The two cases differ principally in the fact that in the former the molecules cross the surface in one direction only, whereas in the latter they cross in both directions. We choose to discuss the latter case.

To be absolutely specific in our thinking, we phrase the question this way: If a particular molecule has crossed a particular element of surface during a particular short interval of time dt, what is the chance that it had the velocity components u, v, w?

This is obviously a case for the application of Bayes'

Theorem; for if we call "having the velocity in question" event A, and "crossing the surface" event B, the question really is, "B having happened, what is the chance that it was accompanied by A?" The answer, we know, is

$$p_B^*(u, v, w) = \frac{p^*(u, v, w)\, p_{uvw}(B)}{p(B)}.$$

Of the quantities on the right, $p^*(u, v, w)$ is given by (73), and of course the denominator can be found by summing expressions such as the one in the numerator, provided $p_{uvw}(B)$ can be found. Virtually the entire problem, therefore, lies in finding this last quantity.

Let us make our choice of axes in such a way that the x-axis is normal to the particular element of area which we have under consideration, and which we will denote by dA. Then whether the molecule in question crosses or does not cross the surface depends upon its position at the beginning of the time interval dt. The simplest method of explaining where it must have been in order to get across is to imagine that it itself is somehow brought to rest at the beginning of the interval, and simultaneously the element of area is given a velocity $-u$, $-v$, $-w$. The relative velocities remain unaltered, and the molecule would have passed through the area if, and only if, the area passes over the molecule. By this process, however, the area is made to generate a solid, the volume of which is its base dA multiplied by its perpendicular height. That height is evidently equal to the arithmetical value of $u\,dt$, which we denote as usual $|u|\,dt$. Hence the volume element is $|u|\,dA\,dt$.

Our molecule, however, was tagged without reference to its position: it is just as likely to be in one part of the vessel as another. If, then, we denote the total volume by V, we have $p_{uvw}(B) = |u|\,dA\,dt/V$. This is our numerator.

To get $p(B)$ we must sum this compound probability over every possible event A with which it may be associated: that is,

over every possible set of values of u, v and w. This gives us, when common factors are cancelled,

$$p_B^*(u, v, w) = \frac{|u| e^{-a(u^2+v^2+w^2)}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |u| e^{-a(u^2+v^2+w^2)}\, du\, dv\, dw}.$$

We have already found that the v and w integrals together give a factor π/a. The u integral is complicated somewhat by the presence of the *absolute value* of u in its integrand. Upon noting, however, that $|u|$ equals $+u$ when u is positive and $-u$ when u is negative, we easily break the integral up into two parts each of which can readily be evaluated. The answer is $1/a$; whence the entire denominator becomes π/a^2.

Thus finally,

$$p_B^*(u, v, w) = \frac{a^2}{\pi} |u| e^{-a(u^2+v^2+w^2)}. \tag{78}$$

This is the answer required.

We shall not follow the problems of the kinetic theory further at this time.[1]

§ 67. *Some Instructive Illustrations; A More Complicated Jacobian*

As a final illustration of the use of the Jacobian in transforming a distribution function from one set of variables to another, we choose the following, which will be of service to us later on.

EXAMPLE 41.—*Six new variables, $\bar{u}$, $\bar{v}$, $\bar{w}$, $\bar{u}'$, $\bar{v}'$, $\bar{w}'$, are defined in terms of six old variables u, v, w, u', v', w', by means of the equations*

$$\left.\begin{aligned} \bar{u} &= u - \lambda S, & \bar{u}' &= u' - \lambda S, \\ \bar{v} &= v - \mu S, & \bar{v}' &= v' - \mu S, \\ \bar{w} &= w - \nu S, & \bar{w}' &= w' - \nu S, \end{aligned}\right\} \tag{79}$$

where S is written briefly for the expression

$$S = \lambda(u - u') + \mu(v - v') + \nu(w - w'),$$

[1] They will be considered in greater detail in Chapter XI.

while λ, μ *and* ν *are constants. What is the ratio of corresponding volume elements* $d\bar{A}$ *and* dA *in the two manifolds?* [1]

By straightforward differentiation we find

$$\frac{\partial \bar{u}}{\partial u} = 1 - \lambda \frac{\partial \mathcal{S}}{\partial u} = 1 - \lambda^2,$$

$$\frac{\partial \bar{u}}{\partial v} = -\lambda \frac{\partial \mathcal{S}}{\partial v} = -\lambda\mu,$$

$$\frac{\partial \bar{u}}{\partial w} = -\lambda \frac{\partial \mathcal{S}}{\partial w} = -\lambda\nu,$$

and so on, the complete Jacobian being

$$\frac{d\bar{A}}{dA} = \begin{vmatrix} 1-\lambda^2 & -\lambda\mu & -\lambda\nu & \lambda^2 & \lambda\mu & \lambda\nu \\ -\lambda\mu & 1-\mu^2 & -\mu\nu & \lambda\mu & \mu^2 & \mu\nu \\ -\lambda\nu & -\mu\nu & 1-\nu^2 & \lambda\nu & \mu\nu & \nu^2 \\ -\lambda^2 & -\lambda\mu & -\lambda\nu & 1+\lambda^2 & \lambda\mu & \lambda\nu \\ -\lambda\mu & -\mu^2 & -\mu\nu & \lambda\mu & 1+\mu^2 & \mu\nu \\ -\lambda\nu & -\mu\nu & -\nu^2 & \lambda\nu & \mu\nu & 1+\nu^2 \end{vmatrix}.$$

This determinant can be greatly simplified by a few elementary transformations. Thus adding the fourth column to the first, the fifth to the second and the last to the third, gives

$$\frac{d\bar{A}}{dA} = \begin{vmatrix} 1 & 0 & 0 & \lambda^2 & \lambda\mu & \lambda\nu \\ 0 & 1 & 0 & \lambda\mu & \mu^2 & \mu\nu \\ 0 & 0 & 1 & \lambda\nu & \mu\nu & \nu^2 \\ 1 & 0 & 0 & 1+\lambda^2 & \lambda\mu & \lambda\nu \\ 0 & 1 & 0 & \lambda\mu & 1+\mu^2 & \mu\nu \\ 0 & 0 & 1 & \lambda\nu & \mu\nu & 1+\nu^2 \end{vmatrix}.$$

[1] "Manifold" is a technical term for what is often called "space of n dimensions." Each of our sets of coordinates would require a six-dimensional space, or manifold, to represent it.

Next, subtracting the first *row* from the fourth, the second from the fifth, and the third from the last gives

$$\frac{d\bar{A}}{dA} = \begin{vmatrix} 1 & 0 & 0 & \lambda^2 & \lambda\mu & \lambda\nu \\ 0 & 1 & 0 & \lambda\mu & \mu^2 & \mu\nu \\ 0 & 0 & 1 & \lambda\nu & \mu\nu & \nu^2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}.$$

As all the elements below the diagonal are now zero, the value of the determinant is simply the product of the terms of the principal diagonal. Hence

$$\frac{d\bar{A}}{dA} = 1.$$

In the case of this transformation, therefore, the elements are equal in both systems.

§ 68. *The General Significance of the Jacobian*

It is not only in probability theory that the Jacobian is of interest. What we have really said in arriving at it is this: We have a function $p(x, y, z, \ldots, w)$, which, when multiplied by $dA = dx\,dy\,dz \ldots dw$ and integrated over a certain part of the manifold in which those variables are represented, gives a number in which we are interested. We also know a set of relations by means of which $x, y, z, \ldots, w$ can be represented in terms of new variables $\xi, \eta, \zeta, \ldots, \omega$. It is therefore a simple matter to substitute these relations in p and get a formula which will give us the value which this function takes at every point of the Greek manifold. However, if we multiply this quantity by $d\alpha = d\xi\,d\eta\,d\zeta \ldots d\omega$, and integrate over the proper part of *this* manifold, we will not get the same number as before, because $d\alpha$ and the dA to which it corresponds are not equal. To get back to the same number, we must multiply p *at every point* by the ratio $dA/d\alpha$ *at that point*, before we integrate.

For example, in computing the area of a circle, which is really

integrating the function $p(x, y) = 1$, we might desire to use polar coordinates r and θ defined by the equations

$$\left.\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta. \end{aligned}\right\} \tag{80}$$

Of course, since p did not contain x and y, it does not contain r and θ either. But it is not true that the area can be found by integrating $dr\,d\theta$ over the circle in question.

In ordinary calculus we say the reason for this is that $dr\,d\theta$ is not the element of area. Instead the element of area is $r\,dr\,d\theta$. It is obvious that this r is nothing else than the Jacobian of the transformation.

In similar fashion, the " element of integration " in the set of coordinates $\xi, \eta, \zeta, \ldots, \omega$ is always

$$\frac{\partial(x, y, z, \ldots, w)}{\partial(\xi, \eta, \zeta, \ldots, \omega)}\,d\alpha.$$

The only difference between the Calculus point of view, which is the more useful one in most lines of investigation, and that which we have used in our discussion of probability, is that in the latter case it proved to be most satisfactory to include the Jacobian as part of the new distribution function, whereas it is generally simpler to think of it as a part of the differential element instead.

PROBLEMS

1. Find the Jacobian of the transformation (80).

2. Find the volume element in cylindrical coordinates.

3. Planck's radiation formula

$$p(\nu) = \frac{a\nu^3}{e^{b\nu} - 1} \tag{81}$$

can be interpreted as a " distribution function " in which the variable is the frequency of the light. That is, the probability of unit energy being emitted between frequencies ν and $\nu + d\nu$ is given by (81). Find the probability of unit energy being emitted between the wave-lengths λ and $\lambda + d\lambda$.

4. What is the probability that the energy of a gas molecule lies between W and $W + dW$?

5. When a gas diffuses outward through an orifice the equilibrium conditions within the enclosure are destroyed. It is therefore no

longer true that the probability of a molecule having specified velocity components is independent of its position. This renders invalid the argument by means of which we determined $p_{uvw}(B)$ in § 66.

Sometimes in physical problems, however, we think of the electrons within a metal as behaving just like the molecules of a gas. If the metal is hot, " thermionic " electrons leak off its surface. They are supposed to get out in such small numbers as not seriously to upset the equilibrium within the metal. Assume this to be true; also assume one is emitted if and only if its x-component of velocity exceeds a positive quantity $\sqrt{E}$.

If the history of an emitted electron is traced back to a moment just before emission, what is the probability that its velocity components were u, v, w?

6. In the case of the thermions mentioned in Problem 5 it is supposed that those which emerge have their v and w unchanged, but that their velocity in the x-direction is changed to a new value u' defined by the law $u^2 - u'^2 = E$.

Assuming this to be true, what is the distribution of velocities after emission?

7. The can containing a gas carries a set of axes x, y, z. It is being translated relatively to a " fixed " set of axes x', y', z' with a velocity U, V, W. By the principle of relativity this is the same thing mechanically as if the x, y, z were " fixed " and the axes x', y', z' were moving with a velocity $-U$, $-V$, $-W$ with respect to them. Of course, if the argument of § 64 is true at all, it applies to the can and the axes x, y, z. Find the chance that a particular gas molecule has components of velocity u', v', w' with respect to the axes x', y', z'.

8. Find the volume element in toroidal[1] coordinates; that is, in a system in which any point P which lies in the xz-plane has the coordinates r, θ, 0, while if it does not lie in that plane, its first two coordinates are determined in exactly the same manner in the plane which contains both the point and the z-axis, while the third coordinate is the angle between this plane and the xz-plane. The system is illustrated by Fig. 21.

9. In aiming at a target, the bull's-eye will not always be hit. We choose to think of firing over so short a range that there is no

---

[1] The name " toroidal " is due to the fact that any surface defined by the equation $r =$ constant is a " torus " (that is, a doughnut). The other coordinate surfaces are: for θ constant, segments of cones having the z-axis as axes; for ϕ constant, planes containing the z-axis.

appreciable curvature of path and make the following assumptions (see Fig. 22):

(*a*) The chance of a horizontal error h in aim is totally independent of any error v vertically.

(*b*) The chance of lying in an angular sector $d\theta$ is the same for any such sector.

Find the chance of a shot having an error between (h, v) and $(h + dh, v + dv)$.

10. In the case of Problem 9, find the chance of the shot hitting between (r, θ) and $(r + dr, \theta + d\theta)$.

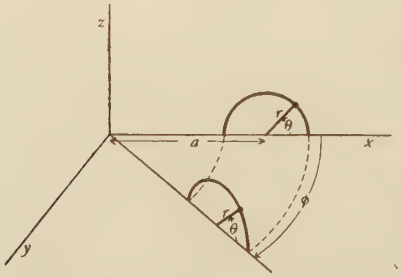


Fig. 21.—Toroidal Coordinates.

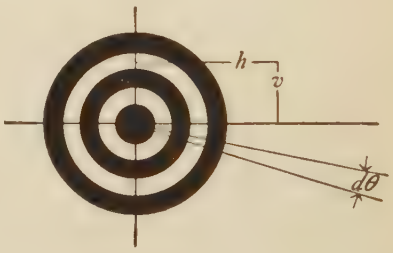


Fig. 22.

11. Suppose the target of Problem 9 to be inclined with respect to the vertical at an angle α. Find the probability of a shot falling between (h', v') and $(h' + dh', v' + dv')$ on this target, the h' and v' being supposed to be measured along its surface.

12. In Problem 9, if we choose any direction θ, there is a distance $r(\theta)$ at which the probability of lying within a given element of area dA is exactly p^*. For some other direction the same will be true, though the new distance need not necessarily be the same. The curve

$$r = r(\theta)$$

along all points of which the probability has the same value p^* is called a "curve of equal probability." Find the equation of these curves.

13. Find the curves of equal probability in Problem 11.

References for Outside Reading

1. Coolidge: *Probability*, Chapters I and V.
2. Czuber: *Wahrscheinlichkeitsrechnung*, pp. 273–286; 389–400.

CHAPTER VII

Averages

§ 69. *Definition of an Average*

The average of a set of quantities is such a quantity that if every member of the set were replaced by it, their aggregate would remain unchanged.

For example, the average weight of a group of men is such that if every man were of the same weight, their aggregate weight would be unaltered. If there were three of them, weighing 140, 160 and 195 pounds, the average would be 165.

Obviously, if the set contains m quantities, of which n_1 have the magnitude a_1, n_2 the magnitude a_2, and so on, this definition is equivalent to the equation

$$m\,\bar{a} = \Sigma\, a_j\, n_j;$$

which leads at once to the formula [1]

$$\mu_1(a) = \bar{a} = \frac{1}{m}\Sigma\, a_j\, n_j \qquad (82)$$

for computing an average.

The following theorem is almost obvious: *If $\bar{a}$ is the average of a set of numbers $a_1, a_2, \ldots, a_n$, and $\bar{b}$ is the average of $b_1, b_2, \ldots, b_n$, the average of the sums $a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n$ is $\bar{a} + \bar{b}$.*

§ 70. *Mathematical Expectation*

Suppose there is a variable a which is capable of taking on any one of the values $a_1, a_2, a_3, \ldots, a_s$, and suppose that

[1] $\mu_1(a)$ means "the first moment of a". The reason for this notation will appear in § 71.

the chance of it taking the value a_1 is p_1, the chance of it taking the value a_2 is p_2, and so on. Suppose finally that a total of m independent trials of this quantity is made, and that in n_1 of them its value is a_1, in n_2 its value is a_2, and so on. Obviously the aggregate value of the quantity a in all the trials is $a_1n_1 + a_2n_2 + \ldots + a_sn_s$, and its average value per trial is

$$\bar{a} = a_1\frac{n_1}{m} + a_2\frac{n_2}{m} + \ldots + a_s\frac{n_s}{m}.$$

We cannot predict in advance what this average $\bar{a}$ is going to be: that is, we cannot predict it with certainty. The only certain way is to try it. But we know from Bernoulli's Theorem that the CHANCE of n_1/m differing from p_1 by an important amount becomes smaller and smaller the larger m is made, and a similar argument applies to the other ratios also. Hence, in a large number of trials there is little chance that $\bar{a}$ will differ much from $a_1p_1 + a_2p_2 + \ldots + a_sp_s$. We call this quantity the "mathematical expectation" of a.

Definition: If a can take only the values $a_1, a_2, \ldots, a_s$ and zero, the probability of each being $p(a_1), p(a_2), \ldots, p(a_s)$ and $p(0)$, the mathematical expectation of a is [1]

$$\epsilon_1(a) = \sum_{j=0}^{s} a_j\, p(a_j). \tag{83}$$

The following theorem is at once obvious:

Theorem: If a large number of independent trials of the value of a are made, the chance that its average value per trial differs from its mathematical expectation by more than some preassigned quantity is small, and may be made as small as we please by sufficiently increasing the number of trials.

Strictly speaking, this theorem is only obvious when the number of possible values of a is finite. For when there is an infinity of terms, it is not necessarily true that the sum of their limits is equal

[1] Read "the expectation of a". The reason for the subscript 1 will appear in § 71. When no confusion can arise, $\epsilon_1(a)$ and $\mu_1(a)$ will often be written simply ϵ_1 and μ_1.

to the limit approached by their sum. In other words, it is not always proper to take limits of individual terms and add them together unless the number of terms in the summation is finite. This is illustrated, for instance, by the set of terms

$$\frac{\sin x}{1}, \quad \frac{\sin 3x}{3}, \quad \frac{\sin 5x}{5}, \quad \ldots,$$

each of which approaches the limit zero as x approaches π. The sum of the limits approached by the separate terms is therefore $0 + 0 + \ldots = 0$. But the sum of the terms themselves is a Fourier series which represents the constant $+\pi/2$ for values of x between 0 and π, and the constant $-\pi/2$ for values of x between π and 2π. Therefore the sum of an infinite set of such terms approaches either $+\pi/2$ or $-\pi/2$ as a limit, according to whether the limit is approached by starting from values of x smaller than π and ascending toward π or by starting from values of x larger than π and descending toward it. In no case, however, can the limit of the sum be made to approach the value 0.

So in the case of our theorem. There are problems in which a very large difference between experimental average and expectation is almost certain, no matter how large the number of trials. One is given in § 78. I know of no case, however, in which such a problem has any practical importance.

Another remark should be made in passing. In the sense in which we are using the words, an "average" is the result of experiment, while a "mathematical expectation" is an advance judgment as to what we had a right to expect that average to be. This usage is not universal. Many writers use the two terms more or less indiscriminately; others use "average" for either idea, and "expectation" only when a valuable consideration is involved.

We shall attempt to keep the ideas distinct; though we shall usually say "expectation" in preference to the more cumbersome "mathematical expectation."[1]

A simple example will make the general idea of expectation somewhat clearer:

EXAMPLE 42.—*Two dice are tossed. If seven appears the player receives a dollar; otherwise nothing. What is his expectation of gain?*

[1] There is another rather mystical concept of "moral expectation" to which we shall refer in passing in § 79. It was introduced years ago in a futile attempt to explain away something which can better be explained without it; and though it is no longer believed in by anyone, it is still accorded pentasyllabic lip-worship in our adherence to the adjective "mathematical."

Here the "quantity a" represents money won, and can take only two values, one and zero. The probability of the first is the same as the probability of a seven appearing, which is $\frac{1}{6}$; the probability of the other is $\frac{5}{6}$. Hence

$$\epsilon_1(a) = \tfrac{1}{6} \cdot 1 + \tfrac{5}{6} \cdot 0 = \tfrac{1}{6}.$$

The player's expectation of gain is therefore one-sixth of a dollar.

Suppose, now, that the player were required to pay a fixed sum A per throw. After a large number of games he would have received some average amount $\bar{a}$ per game: if $A > \bar{a}$ he would be the loser, and if $A < \bar{a}$ the winner. We already know, however, that $\bar{a}$ is not likely to deviate appreciably from ϵ_1 if the number of games is great. Hence if $A > \epsilon_1$, that is, if he pays more than $16\frac{2}{3}$ cents per game, he is almost certain to lose in the long run; for no matter by how little his payments may exceed his expectation, the probability of his average winnings differing from it by as much will approach zero as the number of games is increased. Conversely, if his payments are less than his expectation, he is almost sure to win.

It is obvious, then, that the game is not fair in either of these cases; and we conclude that what he could fairly be required to pay per game is a sum equal to his expectation.

But suppose he did pay $16\frac{2}{3}$ cents per game; then what? After he has played m games he will have won some number n, and his net winnings will have been $n - \frac{1}{6}m$. By the first half of Bernoulli's Theorem we know that if m is large enough n will almost certainly differ from $\frac{1}{6}m$ by more than any pre-assigned quantity, however large. Hence, given games enough, the player is almost sure to *either win or lose* a very large sum; but which it will be we cannot say. By the *second* half of Bernoulli's Theorem we are equally certain that $\frac{n}{m} - \frac{1}{6}$, or in other symbols $\bar{a} - \epsilon_1$, will be negligible: the average net gain *per game* will very probably be small.

Now what is true of this example is true in general: *Whenever a gambling game is conducted by repeated independent trials, a player who pays out more per game than his expectation of winning is virtually certain in the long run to experience an average loss per game substantially equal to the difference; if he pays less than his expectation, he is virtually certain in the long run to experience an average gain per game substantially equal to the difference; while if he pays exactly his expectation his average loss or gain per game will almost certainly be negligible, though the aggregate will probably be large.*

Let us put this in more vivid terms: If you pay too much you lose a lot; if too little the other fellow does; and if just the right amount one of you loses, but not such a whale of a lot.

Insurance viewed from the companies' standpoint nearly duplicates the conditions of this problem: it would exactly duplicate them if every risk were "equally good," [1] equal in amount,[2] and paid for in a single premium.[3]

A few decades ago it was customary for fraternal life companies to pay too much per game (that is, the face value of their policies was higher than the premium justified), and they actually lost money. Conservative companies, on the other hand, pay too little per game (that is, they charge decidedly more than the policies are worth, though the mutual companies return most of it in the form of dividends). They are "gambling on a sure thing." If, instead, they charged the price which is mathematically "fair," they would be about as likely to go bankrupt as not, which would certainly not be a benefit to their policy-holders.

EXAMPLE 43.—*A penny is tossed repeatedly. If heads appears for the first time on the n'th throw, the player receives n cents*

[1] In the case of life insurance, if every man of a given age were equally likely to die at a given time, so far as the company knew. This is not actually true, due to the knowledge they acquire from their physical examinations.

[2] That is, if every life were insured for the same amount.

[3] In practice, the payment of yearly premiums complicates the computation of the company's "expectation of gain." It does not, however, affect what we are about to say.

and a new game begins. Thus if tails appears twice, followed by heads, the three throws constitute a "game," and the "winnings" are three cents. What is the expectation of gain per game?

If heads appears for the first time on the first throw the gain is 1; the probability of this is $p(1) = \frac{1}{2}$. For tails followed by heads the numbers are 2 and $p(2) = \frac{1}{4}$. In general, the probability of a gain j is $p(j) = 1/2^j$. Hence

$$\epsilon_1(a) = \sum_{j=0}^{\infty} \frac{j}{2^j}.$$

This sum can easily be shown to be 2.[1] To make a fair game, therefore, the player would have to pay the "bank" two cents per game.

§ 71. *Derived Averages and Expectations*

From any set of numbers $a_1, a_2, \ldots, a_s$, a host of new sets can be obtained by various arithmetical processes. In particular, the sets $a_1^2, a_2^2, \ldots, a_s^2; a_1^3, a_2^3, \ldots, a_s^3$; and the like, can be built up. Obviously each of these derived sets has its own average, and these derived averages are just as truly descriptive of the *original* set as $\bar{a}$ was. They are the "average square" of a, its "average cube," and so on. We

---

[1] Let $y = x + 2x^2 + 3x^3 + \ldots$. Then $y/x = 1 + 2x + 3x^2 + \ldots$; and $\int \frac{y}{x} dx = a_0 + 1 + x + x^2 + x^3 + \ldots$, the constant of integration being written in the form $a_0 + 1$. But $1 + x + x^2 + \ldots$ is a geometric series and has the sum $1/(1 - x)$. Hence

$$\int \frac{y}{x} dx = a_0 + \frac{1}{1 - x}.$$

$$\frac{y}{x} = \frac{d}{dx}\left(a_0 + \frac{1}{1 - x}\right) = \frac{1}{(1 - x)^2},$$

or

$$y = \frac{x}{(1 - x)^2}.$$

The series in the text is the special case obtained by putting $x = \frac{1}{2}$. It therefore sums up to the value 2.

denote them, either by $\overline{a^2}$, $\overline{a^3}$, . . . , or by $\mu_2(a)$, $\mu_3(a)$, ... Thus in general

$$\overline{a^i} = \mu_i(a) = \frac{1}{m} \Sigma\, a_j^i\, n_j. \tag{84}$$

These quantities are called by the English school of statisticians the "moments" of the set of numbers a_j. The reason for this name is as follows:

Suppose the a's were represented on a horizontal axis, and that a weight n_j were suspended from each point a_j. The dynamical moment of these weights about the point zero would be exactly μ_1; and their "moment of inertia" or second moment would be μ_2. It is a small generalization to speak of an "ith moment" as well.

The sum of the ith powers of the numbers a_1, a_2, . . . would not be changed if each of these *powers* were replaced by $\overline{a^i}$. Of course, replacing every one of the a's by $\sqrt[i]{\overline{a^i}}$ *before* raising them to the ith power would have the same effect. This is called the *root mean*[1] *ith power*. In particular $\sqrt{\overline{a^2}}$ is the "root mean square."

To these "average ith powers" or "ith moments" correspond "expected ith powers" or "ith expectations." Thus, by merely using the definition (83) we obtain

$$\epsilon_i(a) = \Sigma\, a_j^i\, p(a_j). \tag{85}$$

The factor $p(a_j)$ need not be altered, for the probability of a^i having the particular value a_j^i is the same as the probability that a has the value a_j.

§ 72. *Average Values of Continuous Variables*

All of the concepts with which the present chapter deals have exact counterparts in the case of continuous variables. The principal formal difference lies in the substitution of integral signs in place of signs of summation.

To begin with, let us suppose that we are dealing with a variable x, the distribution function for which is known to be

[1] "Average" and "mean" are synonyms.

$p(x)$. Suppose the total range of variation of this variable to be divided up into elementary intervals of length dx; and suppose that a large number of independent trials of x are made. Within any interval $(x_i, x_i + dx)$ will fall n_i of these values, the sum of which may be denoted by s_i. Then it is obvious that

$$n_i x_i < s_i < n_i (x_i + dx).$$

The sum of *all* the values of x is obviously obtained by adding together the sums for the separate intervals. Let it be denoted by s. Then we have

$$\Sigma\, n_i x_i < s < \Sigma\, n_i (x_i + dx).$$

The average value of x, which we denote by $\bar{x}$, is obtained by dividing s by the total number of trials made. Hence we have

$$\Sigma \frac{n_i}{m} x_i < \bar{x} < \Sigma \frac{n_i}{m} x_i + dx\, \Sigma \frac{n_i}{m}. \tag{86}$$

As m becomes infinite the ratio n_i/m is likely to be very nearly equal to the probability of a value falling within the interval $(x_i, x_i + dx)$, which is approximately equal to $p(x_i)\, dx$. Further, if dx is not too large, $\Sigma\, p(x_i)\, x_i\, dx$ does not much differ from $\int p(x)\, x\, dx$, and of course $\Sigma\, n_i/m = 1$. Hence we conclude that the lower bound of the inequality (86) is not greatly different from $\int x p(x)\, dx$, nor the upper bound from this quantity increased by dx. The lower and upper bounds are therefore almost certainly close together, if dx is small; and we conclude that x itself is not likely to deviate much from

$$\epsilon_1(x) = \int p(x)\, x\, dx. \tag{87}$$

This we call the "first expectation of x." Obviously the "ith expectation" is

$$\epsilon_i(x) = \int p(x)\, x^i\, dx. \tag{88}$$

As for the experimental averages themselves, we can say nothing beyond what has already been said: that the chance of their differing much from the corresponding expectations is extremely small, if the number of trials is sufficiently large.

Similarly, if we have any function $f(x)$, the expectation of this function is given by the rule

$$\epsilon_1(f) = \int f(x)\, p(x)\, dx; \tag{89}$$

for by (87)

$$\epsilon_1(f) = \int f\, p(f)\, df,$$

which reduces immediately to (89) because of the rule[1] $p(f) = p(x)/f'(x)$.

In general, however many variables the quantity f may depend upon, its ith expectation is always given by the law

$$\epsilon_i(f) = \int\int\int \cdots \int f^i\, p(x, y, z, \ldots, w)\, dx\, dy\, dz \ldots\ dw. \tag{90}$$

We do not stop to demonstrate that this, too, is a consequence of the definition (87).

As an example, we ask for the "expected energy" in the case of a gas molecule. Energy is $W = \frac{1}{2}ms^2$; hence by (89) and (76)

$$\epsilon_1(W) = \int \tfrac{1}{2}ms^2\, p(s)\, ds$$
$$= 2m\sqrt{\frac{a^3}{\pi}} \int s^4\, e^{-as^2}\, ds.$$

If we set $as^2 = z$, this becomes

$$\epsilon_1(W) = \frac{m}{a\sqrt{\pi}} \int_0^\infty z^{3/2}\, e^{-z}\, dz$$
$$= \frac{m}{a\sqrt{\pi}} \left(\frac{3}{2}\right)! = \frac{3}{4}\frac{m}{a}. \tag{91}$$

[1] To correctly interpret this and similar formulae, the reader must remember that $p(x)\, dx$ means always "the probability that x lies between x and $x + dx$", and $p(f)\, df$ "the probability that f lies between f and $f + df$." Naturally, the *form* of the functions $p(x)$ and $p(f)$ need not be the same.

The result (91) has been obtained from a formula for the speed. It could equally well have been found from (73) directly by the use of formula (90). Thus:

$$W = \frac{m}{2}(u^2 + v^2 + w^2).$$

Hence

$$\epsilon_1(W) = \int\int\int \frac{m}{2}(u^2 + v^2 + w^2)\, p\,(u, v, w)\, du\, dv\, dw$$

$$= \frac{m}{2}\left(\frac{a}{\pi}\right)^{3/2}\left\{\int\int\int u^2\, e^{-a(u^2+v^2+w^2)}\, du\, dv\, dw\right.$$

$$+ \int\int\int v^2\, e^{-a(u^2+v^2+w^2)}\, du\, dv\, dw$$

$$\left. + \int\int\int w^2\, e^{-a(u^2+v^2+w^2)}\, du\, dv\, dw\right\}.$$

The integrations run from $-\infty$ to $+\infty$ in each case.

The first of the three integrals within the braces is easily found to be equal to $\frac{1}{2a}\left(\frac{\pi}{a}\right)^{3/2}$. The others are obviously equal to it because of symmetry. Hence we obtain (91) again for the expected energy.

If we noted the energy of our tagged molecule at a large number of instants far enough apart that they constituted "independent" observations, we could obtain an "average energy," which is nearly certain to be almost equal to $\epsilon_1(W)$. Likewise, if we were to observe all the molecules in the gas at once, they would constitute a large set of "trials." If those trials were independent — that is, if the chance of one having a specified W were independent of what we may have found to be true of the rest[1] — their average energy would probably not differ greatly from $\epsilon_1(W)$. For this reason physicists usually speak of $\epsilon_1(W)$ as the "average energy of the molecules."

[1] They are not independent in fact, for the sum of the energies of all is determined by the principle of the conservation of energy.

§ 73. *The Median*

The median of a set of numbers is that number which occupies the central position when the sequence is arranged in order of magnitude. In other words, there are just as many numbers in the sequence *larger* than the median as there are *smaller* than the median.

Strictly speaking, this definition implies that the number of terms in the sequence is odd. If the number of terms is even, either of the two adjacent numbers, or, better still, their average, can be taken as the median. Due to their arrangement in order of magnitude, it will generally make little difference which of these conventions is adopted.

As an example, the median of the set of numbers

$$-28, -8, -8, -1, -1, +28 \quad +56, +56, +68$$

is -1, since there are just four numbers greater and four numbers less than (or equal to) -1. The average of the sequence, on the other hand, is one-ninth of the sum of the terms, which turns out to be 18.

To the "median" corresponds an "expected median." For suppose the chance of a taking the value a_j is $p(a_j)$. A long series of trials will result in n_1 a_1's, n_2 a_2's, and so on. Let these be arranged in order, and call the middle one a_i. Then, obviously

$$n_1 + n_2 + \ldots + n_{i-1} < \frac{m}{2},$$

$$n_1 + n_2 + \ldots + n_i \quad > \frac{m}{2}.$$

But in the long run n_j/m is not likely to differ much from $p(a_j)$; so if we divide these inequalities by m throughout they immediately suggest the relations

$$p(a_1) + p(a_2) + \ldots + p(a_{i-1}) < \tfrac{1}{2},$$
$$p(a_1) + p(a_2) + \ldots + p(a_i) \quad > \tfrac{1}{2}.$$

This is our definition of the expected median.

Hence: *The expected median a_1 is such a number that there is less than an even chance of either a or a_1 being greater than the other.*

For example, in tossing two dice the chances of various sums are as follows:

Sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The expected median is therefore 7, for the chance of the sum exceeding 7 is $\frac{15}{36}$, and the chance of 7 exceeding the sum is also $\frac{15}{36}$.

The general idea is that of a number which we can expect to be exceeded as often as not; but its own chance of repeated occurrence spoils so simple a definition when the variable is capable of only a discrete set of values. On the other hand, if the variable is capable of continuous variation this definition is strictly true, for in that case the chance of the variable taking *exactly* its median value is zero. By a line of argument exactly similar to that used in § 72 we can show that the expected median of x is that value x_m for which

$$\int_{x_m}^{\infty} p(x)\,dx = \tfrac{1}{2}.$$

Of course the integral of $p(x)$ between the limits $-\infty$ and x_m is also $\frac{1}{2}$.

§ 74. *The Deviation*

So far we have introduced three fundamental probability concepts: that of the "most probable" result, of the "expected" result, and of the "expected median"; and have intimated that a host of derived, or secondary, concepts are possible, of which "expected ith powers" have been specifically mentioned. These derived concepts are seldom used in practical studies, however, except in connection with the "deviations" of a set of numbers from the average of the set.

Thus, to the set of numbers $a_1, a_2, \ldots, a_s$ repeated $n_1, n_2, \ldots, n_s$ times, respectively, there corresponds an average $\bar{a}$ and a set of deviations from that average

$$\begin{aligned} d_1 &= a_1 - \bar{a}, \\ d_2 &= a_2 - \bar{a}, \\ &\cdots \qquad \cdots \\ d_s &= a_s - \bar{a}, \end{aligned} \tag{92}$$

each of which recurs as often as the corresponding a. Obviously the set of d's has its own average, median, and the like.

Of these two are especially important:

The average of a set of d's is zero.

By definition $\bar{d}$, or $\mu_1(d)$, is

$$\mu_1(d) = \frac{1}{m} \Sigma d_j n_j.$$

But from (92)

$$\Sigma d_j n_j = \Sigma a_j n_j - m\bar{a};$$

and by (82) $\Sigma a_j n_j = m\bar{a}$. Hence

$$\mu_1(d) = 0. \tag{93}$$

The mean square deviation of a set of numbers is equal to the mean square of the set diminished by the square of their average.

For

$$\begin{aligned} \mu_2(d) &= \frac{1}{m} \Sigma n_j (a_j - \bar{a})^2 \\ &= \frac{1}{m} \Sigma n_j a_j^2 - \frac{2\bar{a}}{m} \Sigma n_j a_j + \bar{a}^2. \end{aligned}$$

But by the use of (82) and (84) this reduces immediately to[1]

$$\mu_2(d) = \overline{a^2} - \bar{a}^2. \tag{94}$$

[1] This equation could obviously be written in either of the forms

$$\overline{d^2} = \overline{a^2} - \bar{a}^2 \quad \text{or} \quad \mu_2(d) = \mu_2(a) - [\mu_1(a)]^2.$$

Similarly (93) could be written $\bar{d} = 0$.

The square root of this quantity μ_2, which we would naturally call the "root mean square deviation," is generally referred to in statistical works as the "standard deviation."

Finally, if a set of numbers $a_1, a_2, \ldots, a_s$ have an expectation $\epsilon_1(a)$, the quantities

$$\begin{aligned} \delta_1 &= a_1 - \epsilon_1, \\ \delta_2 &= a_2 - \epsilon_1, \\ &\ldots \qquad \ldots \\ \delta_s &= a_s - \epsilon_1 \end{aligned} \tag{95}$$

are their deviations from expectation. The probabilities of these deviations being the same as the probabilities of the original a's, it follows that

$$\begin{aligned} \epsilon_1(\delta) &= \Sigma\, p(a_j)(a_j - \epsilon_1) \\ &= \Sigma\, a_j\, p(a_j) - \epsilon_1 \Sigma\, p(a_j) \\ &= \epsilon_1 - \epsilon_1 \Sigma\, p(a_j). \end{aligned}$$

As the summation covers every possible value of a_j, including zero, the set of a's is complete, and $\Sigma\, p(a_j) = 1$. Hence

$$\epsilon_1(\delta) = 0. \tag{96}$$

This corresponds to (93).

Similarly

$$\begin{aligned} \epsilon_2(\delta) &= \Sigma\, p(a_j)(a_j - \epsilon_1)^2 \\ &= \Sigma\, a_j^2\, p(a_j) - 2\epsilon_1 \Sigma\, a_j\, p(a_j) + \epsilon_1^2 \Sigma\, p(a_j) \\ &= \epsilon_2(a) - [\epsilon_1(a)]^2. \end{aligned} \tag{97}$$

This corresponds to (94).

It is obvious that all the formulæ of this section apply to continuous variables as well as to those which are capable of taking only a discrete set of values.

§ 75. *Résumé*

Most of the concepts which we have introduced in this chapter deal with quantities which may take various magni-

tudes. This is, in a way, a limitation of the general idea of an "event" and its probability, for not every aspect of an "event" is numerically measurable. We now frankly limit outselves to such things as are measurable, however, and speak for the future only of the probability of variables taking various values. What we have so far learned then takes the following form:

Concepts Associated with the Results of Experiment	*Concepts Associated with Advance Judgment*
(1) [Of a finite set of numbers, the one which occurs most frequently is called the "mode"].[1]	(1) Of a finite set of numbers, one is the "most probable." (It is also called the "mode" by some writers.)
(2) If the numbers are arranged in order of magnitude, the one which occupies the middle position is the "median."	(2) There is a value which the variable has at most an even chance of exceeding, and at least an even chance of either equalling or exceeding. It is the "expected median."
(3) That number by which every number of the set could be replaced without changing their sum is the "average" or "mean" of the set.	(3) There is an average per trial which a number is most likely to show after m independent trials have been made. This average approaches a limiting value as m is indefinitely increased. This limiting value is the "expectation" of the number.
(4) If each number is reduced by an amount equal to the average of the set, the resultant numbers are the "deviations from the average."	(4) If every possible value is reduced by an amount equal to the expectation of the set the results are the "deviations from expectation."
(5) The "mean deviation" is zero.	(5) The "expected deviation" is zero.
(6) If each deviation of the set is raised to the ith power, the average of the result is called "the ith moment" of the deviations.	(6) If every possible deviation from expectation is raised to the ith power, the resultant set of numbers possesses its own expectation. This is called the "ith expectation" of the deviations.

[1] The statement in brackets is not made in the text.

§ 76. *Some Instructive Illustrations; The General Case of Independent Trials*

Let us find the expected number of successes in m independent trials of an event, when the probability of its occurrence on a single trial is p.

The formula for the probability of n successes is

$$p(n) = C_n^m \, p^n \, (1 - p)^{m-n}.$$

As the quantity in which we are interested is the number of occurrences of the event, $a_n = n$. The expectation of n then becomes

$$\epsilon_1(n) = \sum_{n=0}^{m} n \, p(n)$$

$$= \sum_{n=0}^{m} n \, C_n^m \, p^n \, (1 - p)^{m-n}.$$

But

$$n \, C_n^m = m \, C_{n-1}^{m-1},$$

and

$$m - n = (m - 1) - (n - 1).$$

Hence

$$\epsilon_1(n) = mp \sum_{n=0}^{m} C_{n-1}^{m-1} \, p^{n-1} \, (1 - p)^{(m-1)-(n-1)}.$$

This summation, however, is of exactly the form [1] (15), except that m is replaced by $m - 1$ and n by $n - 1$. Hence the sum of all the terms is $[p + (1 - p)]^{m-1} = 1$. It therefore follows that $\epsilon_1(n) = mp$. *In this case the expected number of successes and the most probable number come out to be equal*, except that the expected number may be fractional, whereas the most probable number is necessarily an integer (see § 38).

---

[1] There is one term too many, but this term turns out to be zero because of the factor C_{-1}^{m-1}, which vanishes.

§ 77. *Some Instructive Illustrations; The General Case of Dependent Trials of the Sort Discussed in § 27*

The discussion in § 27 concerned itself with the probability of drawing just p red and q black balls from an urn which contained m red and n black balls, assuming that after a ball had been withdrawn it was not replaced before the next drawing was made. The general result was given by equation (25):

$$P_{m,n}(p, q) = \frac{C_p^m C_q^n}{C_{p+q}^{m+n}}. \tag{25}$$

It is desired to find the number of red balls which may be expected to be drawn in $p + q = r$ trials.

In this case the quantity under discussion — the number of red balls — is obviously measured by p itself. In other words, $a_p = p$. Moreover, since $p + q$ is equal to the constant r, q is also a function of p and must be replaced by its value $r - p$. Thus we find

$$\epsilon_1(p) = \sum_{p=0}^{r} p \frac{C_p^m C_{r-p}^n}{C_r^{m+n}}.$$

Noting that C_r^{m+n} does not vary from term to term of the summation; and replacing $p\,C_p^m$ by $m\,C_{p-1}^{m-1}$, this becomes

$$\epsilon_1(p) = \frac{m}{C_r^{m+n}} \sum_{p=0}^{r} C_{p-1}^{m-1} C_{r-p}^n.$$

The sum of binomial coefficients contained in this equation is in the form (26), as can easily be seen by writing $C_{(r-1)-(p-1)}^n$ in place of C_{r-p}^n. Hence

$$\epsilon_1(p) = m \frac{C_{r-1}^{m+n-1}}{C_r^{m+n}},$$

which works out to be $\dfrac{mr}{m+n}$.

As in the last example, when this result is integral, it is also the most probable number of red balls, and when it is fractional the most probable number of red balls is one of the adjacent integers.

§ 78. *Some Instructive Illustrations; A Dice Problem*

In the illustrations considered in the last two sections the "expectation" was either equal to or adjacent to the "most probable" result, according to whether both results were integral or not. The following example serves to show that this condition does not always exist.

EXAMPLE 44.—*A die is tossed until an ace appears. What is the most probable number of throws, and what is the expected number of throws?*

The chance that the ace appears on the first throw is $p_1 = \frac{1}{6}$. The chance that it does not appear on the first throw but appears on the second is $p_2 = \frac{5}{6} \cdot \frac{1}{6}$. The chance that it does not appear on either of these throws but appears on the third is $p_3 = (\frac{5}{6})^2 \cdot \frac{1}{6}$, and in general, the chance that it appears for the first time on the jth throw is $p_j = (\frac{5}{6})^{j-1} \cdot \frac{1}{6}$. It is obvious that the largest of these probabilities is p_1. Hence the most probable number of throws is 1.

The expected number of throws, on the other hand, is $\sum_{j=1}^{\infty} j\, p_j$, or

$$\epsilon_1(j) = \tfrac{1}{6}[1 \cdot 1 + 2 \cdot \tfrac{5}{6} + 3 \cdot (\tfrac{5}{6})^2 + 4 \cdot (\tfrac{5}{6})^3 + \ldots].$$

But from the footnote at the end of § 70

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \ldots$$

Comparing this with $\epsilon_1(j)$ we easily get the result

$$\epsilon_1(j) = 6.$$

In other words, the expected number of throws is six, though the most probable number is one.

§ 79. *Some Instructive Illustrations; The St. Petersburg Paradox*

As a final illustration of this sort of computation we take the following very famous problem, the apparently absurd solution of which has puzzled men for generations:

EXAMPLE 45.—*A penny is tossed until heads appears. If heads appears on the first throw the bank pays the player one dollar. If heads*

appears for the first time on the second throw the player receives two dollars. If heads appears for the first time on the third throw he receives four dollars. If it appears for the first time on the fourth throw he receives eight dollars, and so on. What should the player pay the bank for the privilege of playing a sequence of this sort in order that the game may be equitable?

The probability of heads appearing for the first time on the nth throw is $p_n = (\frac{1}{2})^n$. If so the player receives an amount $a_n = 2^{n-1}$. Substituting these values in (83) it becomes

$$\epsilon_1(a) = \Sigma\, p_n a_n = \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \ldots .$$

This sum is to be extended to every possible value of n. However, there is no logical limit to the number of tails which may appear before the first head shows up. It is possible for heads to appear first on the millionth throw, although, of course, the probability of any such thing occurring is extraordinarily small. This means, of course, that an infinity of $\frac{1}{2}$'s must be added together, with the consequence that $\epsilon_1(a) = \infty$. In other words, in order to play a sequence of this sort fairly to the bank, the player must first deposit with the bank an infinite amount of money.

From a common sense standpoint this result is absurd. No sane man would ever consider paying the bank one hundred dollars for such a chance, much less an infinite amount. And yet the mathematics itself is straightforward. It is certainly no more questionable than the mathematical processes used throughout the remainder of this book. If, therefore, the result is incorrect, it throws suspicion upon the entire structure of Probability Theory. It is therefore essential to know why the result does not agree with common sense.

A number of answers have been given to this question. Probably the first historically is that given by Daniel Bernoulli, who distinguished between what he termed "mathematical" and "moral" expectation. The former he defined exactly as above, but differentiated the latter from it by including certain psychological considerations of the following sort:

A dollar is worth more to a beggar than it is to a millionaire. In fact, says Bernoulli, the satisfaction which one gets out

of any acquired sum of money is less and less, the greater the amount of money which he already has. When, therefore, in the game under consideration, the player pays out from his moderate fortune a certain sum of money, he pays out something the *moral* value of which is comparatively large. He has a chance — though a very small chance — of winning an enormous amount of money, provided heads is long delayed. Suppose he is fortunate and does win this enormous amount of money. He is then a very wealthy man, and his winnings acquire a *moral* value based upon his new, rather than upon his old, economic standing. In other words, his losses, being based upon a comparatively low economic standing, loom larger in his estimation than do his winnings.

By introducing the law that the relationship between the moral and mathematical expectations is of a logarithmic nature, the computation of the amount which he should pay can be carried out and serves to give a result which is somewhere within the bounds of reason. It possesses its own common-sense absurdity, however, in the fact that the amount which the player should pay to the bank depends upon the player's own economic standing as well as upon what he can expect to win, and is therefore different for *different men*, even though they play against the *same bank*. What the banker would say to such an arrangement is quite obvious. As this is certainly not the true way out of the dilemma it is not necessary to go into the mathematics of it.

Another explanation which has been favored by many statisticians is based upon the fact that, if the problem is slightly modified, it gives a result which is not quite so absurd. To see this, suppose that instead of using the set of values $a_1 = 1,\ a_2 = 2,\ a_3 = 4, \ldots,\ a_n = 2^{n-1}$, the set were taken as $a_1 = 1,\ a_2 = x,\ a_3 = x^2, \ldots,\ a_n = x^{n-1}$, where $x < 2$. Substituting these values (along with the values of p_j, which remain unchanged) in (83) it takes the form

$$\epsilon_1(a) = \frac{1}{2} + \frac{1}{2}\cdot\frac{x}{2} + \frac{1}{2}\cdot\left(\frac{x}{2}\right)^2 + \ldots = \frac{1}{2 - x}.$$

This expression is finite. For instance, if x is 1.5 instead of 2, the player should pay the bank two dollars for the privilege of playing a sequence, which seems reasonable. However, as x approaches the value 2 the sum gets larger and larger, and from the mathematical standpoint ultimately becomes infinite.

The second explanation of the paradox makes use of these facts, and then concludes that we have no intuitive sense of the immensity of the difference between the expectation when $x < 2$, and the expectation when $x = 2$, and therefore are unwilling to pay the amount of money which is logically called for. This explanation, too, is open to objection, however; for if x were taken as 1.99 instead of 2 (so that instead of paying \$2.00 if heads appears on the second throw the bank would pay only \$1.99) the amount which the player should pay the bank works out to be \$100, and no sensible man could be induced to pay this amount for playing the modified game. Hence, if the trouble is with our intuition, that intuition must go wrong even when dealing with numbers which deviate from \$2.00 by amounts which the department stores have done their utmost to make familiar to us.

I believe the true explanation of the paradox is quite different from either of these, and is based upon the fact that in our every-day experience we have to deal only with individuals who have finite fortunes and who would therefore be incapable of paying back the sums which are required in those very rare cases where heads appears only after an extremely long run of tails. To see what the effect upon the mathematical expectation is, if the bank has limited wealth, consider the following alternative form of the problem:

EXAMPLE 46.—*What is the equitable payment for playing in the game described in Example 45, if the bank's wealth is limited to \$1,000,000?*

The probability of heads appearing first on the nth throw is $(\frac{1}{2})^n$, as before. If so, the bank pays $\$2^{n-1}$ if this is less

than $1,000,000, otherwise it pays $1,000,000. In other words,

$$\left.\begin{aligned} p_n &= \frac{1}{2^n}, \\ a_n &= 2^{n-1}, \end{aligned}\right\} \quad \text{if } 2^{n-1} < 1{,}000{,}000;$$

$$\left.\begin{aligned} p_n &= \frac{1}{2^n}, \\ a_n &= 1{,}000{,}000, \end{aligned}\right\} \quad \text{if } 2^{n-1} > 1{,}000{,}000.$$

Now $2^{19} < 1{,}000{,}000$ and $2^{20} > 1{,}000{,}000$. Hence the first set of conditions applies for $n \leqq 20$; the second set for $n > 20$. Thus (83) becomes:

$$\epsilon_1(a) = \sum_{n=1}^{20} (\tfrac{1}{2})^n 2^{n-1} + \sum_{n=21}^{\infty} (\tfrac{1}{2})^n 1{,}000{,}000.$$

The first summation obviously gives 10. The second is a geometric series, of which the sum is

$$\frac{1{,}000{,}000}{2^{20}} = 0.9536.$$

In order for play to be equitable against a million-dollar bank, therefore, the player should pay $10.95 — a reasonable amount.

If the bank had a billion dollars, the payment would be less than $16; while if the bank had $1,000,000,000,000 — probably more than the total economic value of the earth — the payment would be less than $21.

Taking the other extreme, if the bank had only $100 capital,[1] the payment would be $4.26. If it had $10, it would be $2.63, while if it had only $1 (in which case the payment would be $1, no matter how many tails appeared before the first head), the payment would be $1, as it should.

[1] All these numbers are based upon the bank's wealth after receiving the payment, not before. In the case of the million-dollar bank the difference is negligible; in some of the figures which follow, however, it is not.

I believe this to be the true explanation of the paradox. If the bank were infinitely wealthy, the expectation would be infinite. Therefore the mathematics is correct in either case. But we are accustomed to deal only with "limited wealth" and cannot conceive of "unlimited amounts of money."

In other words, our intuition is in error, rather than the theory; but only because the theory deals with material for which we have had no opportunity to build up intuitive judgments.

§ 80. *The Expectation of a Probability*

The concept of an expectation can be applied to any number which may be determined from the result of an experiment. Among the things which may be determined about an experiment is, however, the *a priori* probability of the exact result which has been obtained. For example, if we toss two dice we determine a certain sum from 2 to 12, according to the faces which happen to appear, each of which sums has a definite probability of appearance, as we saw in Problem 3, § 47. The exact relationship is

Sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

If, then, our experiment gives us a sum 4, we have a result of which the *a priori* probability was $\frac{3}{36}$; while if it gives a sum 12 we have something of which the *a priori* probability was only $\frac{1}{36}$.

These numbers measure the unusualness of our result: if they are very small, our result is extraordinary, while if they are large it is not at all so. If we were to make a large number of trials, we could find the "average unusualness" of our results by averaging the *a priori* probabilities of the results obtained.

Similarly, we may compute the "expectation of the probability," just as we may compute the expectation of any other number, and thus find a sort of theoretical measure of the degree to which the result of an experiment may be expected to be unusual.

Let us take as an example the numbers given above. The chance of getting a sum 2 is $\frac{1}{36}$; if we do, the number with which we are dealing—the probability—is also $\frac{1}{36}$. So this event contributes a term $(\frac{1}{36})^2$ to our expectation. Similarly the chance of getting a sum 3 is $\frac{2}{36}$, which contributes a term $(\frac{2}{36})^2$ to our expectation. In

general, any event which has the probability p contributes a term p^2 to the expectation of p; for the number of which we are computing the expectation is p, and its probability of occurrence is also p. In our particular illustration the sum is therefore

$$(\tfrac{1}{36})^2 + (\tfrac{2}{36})^2 + (\tfrac{3}{36})^2 + \ldots + (\tfrac{2}{36})^2 + (\tfrac{1}{36})^2 = 0.1126.$$

In other words, we may expect the experiment to give a result of which the probability is a little greater than $\frac{1}{9}$.

We shall now prove that *the expectation of the probability of a complete set of events is least when the events are equally likely.*

Suppose the events are s in number, and that their respective probabilities are $p_1, p_2, \ldots, p_s$. Then the expectation of their probability is

$$\epsilon(p) = p_1^2 + p_2^2 + \ldots + p_s^2,$$

where the p's must satisfy the condition

$$p_1 + p_2 + \ldots + p_s = 1,$$

since the set is complete. Solving the latter of these equations for p_s and substituting the result in the other, we get

$$\epsilon(p) = p_1^2 + p_2^2 + \ldots + p_{s-1}^2 + (1 - p_1 - p_2 - \ldots - p_{s-1})^2.$$

In this equation the variables are all independent. We may therefore find the minimum value of ϵ by differentiation in the usual way. Differentiating with respect to each of the variables in turn we get expressions of the form

$$\frac{\partial \epsilon}{\partial p_i} = 2p_i - 2(1 - p_1 - p_2 - \ldots - p_{s-1}) = 2(p_i - p_s).$$

In order that all these equations may be zero it is necessary that each p_i be equal to p_s: that is, that every p must be equal. This proves the theorem.

As an illustration, the eleven sums that may be obtained in tossing two dice are not equally likely. We have seen that the expectation of their probability is 0.1126, which, according to the theorem, should be bigger than if they were equally likely. If they were all equally likely, however, the probability of each would be $\frac{1}{11}$; and since p could take no other value than this, its expectation must also be $\frac{1}{11}$. This is indeed less than 0.1126.

A similar theorem can be proved for the case of continuously distributed variables. It reads: *The expectation of $p(x)$ is least if x is distributed at random.* We shall not stop to prove it.

Instead, we may observe that the sort of computation which we have just carried out for p could equally well be carried out for some function of p; for any experiment which fixes a value of p also fixes the value of any function of p. Later on, in our discussion of the Kinetic Theory of Gases, we shall find that the expectation of the *logarithm* of p plays an important part, and in anticipation of our needs in that connection we may prove the following variation upon the theorem which we have just stated:

The expectation of $\log\ p(x)$ is a minimum if the variable x is distributed at random.

For this proof we make use of (89), the function $f(x)$ being for our present purposes $\log\ p(x)$. Our problem, then, is to make

$$\epsilon[\log p(x)] = \int p(x) \log p(x)\, dx$$

a minimum, subject to the condition that

$$\int p(x)\, dx = 1.$$

This is a problem in the Calculus of Variations, and specifically a problem of what is known as the "isoperimetric type." It would carry us too far afield to attempt to explain the theory that underlies such problems, so we shall content ourselves with giving a categorical statement of the rule to be followed in their solution, and showing that the application of that rule leads to the result stated in our theorem.

The rule is as follows: If it is desired to find a function $p(x)$ which will make the integral

$$F = \int f(p)\, dx$$

a minimum, subject to the condition that the integral

$$G = \int g(p)\, dx$$

must always be kept constant, it is only necessary to make the quantity

$$F - \lambda G = \int [f(p) - \lambda g(p)]\, dx$$

a minimum *without any restrictive conditions whatever*. The λ is a constant the value of which can usually be found after the form of $p(x)$ has been determined. Applied to our problem, this rule requires that we make the integral

$$I = \int p\, [\log p - \lambda]\, dx$$

a minimum.

Suppose, now, that we had somehow found the solution to the problem, and that it was

$$p = P(x).$$

Since this function makes I a minimum, it follows that any change we might make in the form of $P(x)$ must of necessity increase I. For example, if we made a small change to the new function

$$p = P(x) + \delta(x)$$

this would have to be true. But if we substitute this result in I we obtain

$$\int [P + \delta]\, [\log (P + \delta) - \lambda]\, dx,$$

which is approximately equal to

$$\int P\, (\log P - \lambda)\, dx + \int (\log P + 1 - \lambda)\, \delta\, dx + \frac{1}{2} \int \frac{1}{P}\, \delta^2\, dx + \ldots,$$

since δ is very small.

Now, it is easy to show that unless the quantity by which δ is multiplied in the second integral is zero we can make the entire result smaller than the first term; which is absurd, since the first term is a minimum by hypothesis. To see this, we remember that $\delta(x)$ is a purely arbitrary function of x. If we were to choose it so that it is negative wherever $\log P + 1 - \lambda$ is positive, and *vice versa*, the second integral would obviously be negative. And if we were to choose

it small enough, the remaining terms would be negligible. The sum of all these terms would then be negative. As this is impossible by hypothesis, it follows that

$$\log P = \lambda - 1,$$

which is a constant. However, if $\log P(x)$ is a constant, $P(x)$ must also be. That is, the variable x must be distributed at random, which proves our theorem.

PROBLEMS

1. In § 76, $\epsilon_1(n)$ was found for the case of *independent* trials. By transformations which are exactly similar to those used there, the summations defining $\epsilon_2(n)$, $\epsilon_3(n)$, . . . can be reduced to one or more terms of the form (15). Find $\epsilon_2(n)$ and $\epsilon_3(n)$.

2. Find $\epsilon_1(\delta)$, $\epsilon_2(\delta)$, $\epsilon_3(\delta)$.

3. Toss a penny ten times and record the number of heads appearing. Call it n. Repeat the experiment until you have 50 values of n. With these *experimental* results find their average, the set of deviations d, and the three moments $\mu_1(d)$, $\mu_2(d)$, $\mu_3(d)$. (It will save time to take ten coins and toss them at once. The number of heads can then be counted after each throw.)

4. Find the " expectation of n," and the first three expectations of δ. The results of Problem 2 will aid you.

5. Suppose the game described in Example 45, § 79, is altered so that if heads do not appear within ten throws the bank captures the stakes, and a new game begins. What is the player's expectation of gain?

6. Equation (77) is the " Normal Probability Law." Find the " most probable velocity " in the x-direction.

7. Find σ, the " standard deviation " of u.

8. Suppose σ were adopted as the unit of velocity, and denote velocities measured in this new system by u'. Find $p(u')$. We refer, of course, to the velocities in the x-direction only.

9. Find the first three expectations of δ under the conditions of Problem 8. You should not need to compute the first two.

10. In Example 45 the bank has $1,000,000; but instead of paying 2^{n-1} dollars if heads appears first on the nth throw, it pays $(1.99)^{n-1}$. What is a fair price per game? What is the fair price if the bank pays $(1.5)^{n-1}$ dollars? Compare these values with the results obtained in § 79 upon the assumption that the bank's wealth was infinite.

11. Ten dice are tossed together, the experiment being repeated fifty times. What is the expectation of the number of times three aces appear?

12. If an experiment produces two numbers a and b, and if the value of a which appears is independent of the value of b, show that the expectation of their product is the product of their expectations.

13. The face cards are discarded from two packs, and thereafter one card is dealt from each. What is the expectation of the product of the numbers appearing on them?

14. If an experiment produces two numbers a and b, show that $\epsilon_1(a + b) = \epsilon_1(a) + \epsilon_1(b)$.

15. The face cards are discarded from a single pack, and then two cards are drawn. What is the expectation of the sum of the numbers appearing on them?

16. Show that the expectation of the probability of a continuous variable is least when the variable is distributed at random.

References for Outside Reading

1. Coolidge: *Probability*, pp. 52–73.
2. Arne Fisher: *Mathematical Theory of Probabilities*, pp. 49–53.
3. Czuber: *Wahrscheinlichkeitsrechnung*, pp. 72–79, 142–146, 239–261.
4. Whitworth: *Choice and Chance*, Chapters IX, X and XI.

CHAPTER VIII

The Distribution Functions Most Frequently Used in Engineering

§ 81. *Introductory*

For many years the only distribution function which scientists were accustomed to use was the Gaussian or Normal Law. This, when written in one variable, as was usually the case, was exactly the equation (77). Various derivations of it have been given; but none of them is satisfactory in a practical sense, for they are all based upon assumptions of such a nature that it is quite impossible to judge whether or not they are approximated in any actual case. For example, it is very frequently assumed that the deviation of any given result from expectation is due to the superposition of a very large number of contributory causes, no one of which is comparable in magnitude to the combined effects of all the rest, and which are just as likely to produce positive as negative deviations. It is generally hopeless to attempt to justify the use of a formula based upon such hypotheses as these, for we do not ordinarily have any clear-cut picture of the causes of deviation to begin with, either as to number or as to the matter of their tendency to produce opposite effects with equal frequency.

In addition to this purely theoretical objection to such proofs, there is the further objection that experience has taught that very few sets of experimental data appear to follow the law — in fact, very few of them are even symmetrical. Hence, though it can do us no harm to know how far men have been able to go in putting a foundation under an ancient monument, we shall be wise not to overlook the fact that it

is, in a sense at least, venerable principally for its age. It has its uses, but it is not divinely ordained for the cure of all statistical woes.

Other distribution functions which have to do with continuously varying quantities are in no better logical standing. It is only in the case of quantities which, in the nature of things, can take only discrete values (such, for example, as the number of objects possessing a given property) that anything approaching practically applicable hypotheses have been found.

It is the purpose of the present chapter to segregate these two classes of distribution functions, and so far as is possible without too tedious theorizing to show under what conditions the individual functions may be expected to be applicable. They may then serve as guides in the discussion of such data as is actually met in practice, especially in cases where the attempt to find the true law by an independent investigation appears to be entirely hopeless. In other cases such an attempt should undoubtedly be made, provided the problem is important enough to justify the labor. It is the only sure way.

§ 82. *Distribution Functions for Discrete Variables; The Binomial Law and Various Approximations to It*

The first law of any general consequence which we met in the course of our studies was

$$P_m(n) = C_n^m p^n (1 - p)^{m-n}. \tag{23}$$

As we know, it represents the probability of n successes in m *independent* trials, if the chance of success in a single trial is p.

This law is exact, and not only that, we know pretty well what the conditions underlying it are. It is true, I suppose, that there are comparatively few practical situations in which the same essential conditions can be maintained for a great

length of time; but there are many problems in which conditions approximate stability to such an extent that we feel no hesitancy in dealing with them on this basis. For example, take the production of stamped parts made on a punch press. The die in use is certain to wear, and thus produce a progressive tendency of some sort; but the trend will usually be slow enough that, over a sufficiently limited portion of the life of the die, this trend may be ignored. So, too, with many other features of the process: sheets differ somewhat in thickness, temperatures change, and so on to a great number of factors. Yet if we sort the product into two sorts, " bad " and " good," the various pieces have something like the same chance of being good.

The Binomial Law, therefore, is one of very broad utility. Its chief objectionable feature is the difficulty of computing it, particularly when the answer desired is the probability of *exceeding* n instead of *equalling* it, in which case a large number of terms might have to be calculated and added together, if m is large. There are, however, fairly accurate approximations which can be used in such cases, the foundations for deriving which have been laid in §§ 42 and 43. We shall now complete the proof.

To get a mental picture of what the proof is to contain we refer back once more to the discussion of Bernoulli's Theorem, and in particular to the accompanying Figs. 8 and 9. These figures were drawn for the particular value $p = \frac{1}{3}$, and present the distribution functions for n successes in exactly m trials for several values of m. We saw that as the number of trials is increased the distribution function becomes flatter and flatter and spreads out more and more widely along the n-axis (Fig. 8), but that when plotted against n/m as in Fig. 9 it becomes higher and spreads less widely as m increases. If we carry this process one step further and plot the distribution function in terms of $n/\sqrt{m}$, we obtain the set of curves shown in Fig. 23. It is obvious that they are very similar to one another, and seem to approach a definite limit as m becomes infinite. It ought to be possible, therefore,

to find some sort of smooth curve which would be a good approximation, at least when m is very large. It should be noted, however, that in order to get the degree of similarity exhibited by Fig. 23 it has been necessary to shift the curves

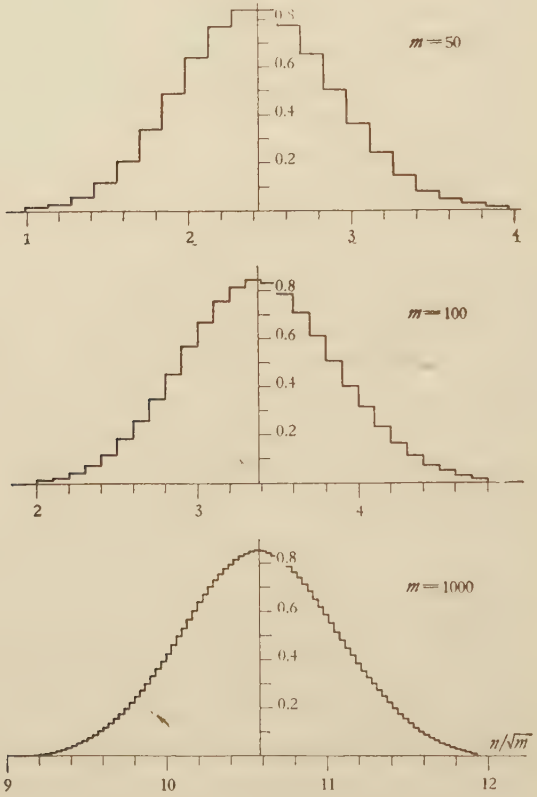


FIG. 23.—AN ALTERNATIVE FORM OF FIG. 7.

so as to cause their maxima to fall vertically above one another.

We begin by replacing all the factorials in (23) by their Stirling approximations. The result is

$$P_m(n) = \frac{1}{\sqrt{2\pi mp(1-p)}} \left(\frac{mp}{n}\right)^{n+\frac{1}{2}} \left(\frac{m-mp}{m-n}\right)^{m-n+\frac{1}{2}} f(m, n), \quad (98)$$

where $f(m, n)$ is written briefly for the series terms

$$\frac{\left(1+\frac{1}{12m}+\frac{1}{288m^2}+\cdots\right)}{\left(1+\frac{1}{12n}+\frac{1}{288n^2}+\cdots\right)\left(1+\frac{1}{12(m-n)}+\frac{1}{288(m-n)^2}+\cdots\right)}$$

which come from (41). It is equivalent to

$$f(m, n) = 1 - \frac{1}{12}\frac{m^2 - mn + n^2}{mn(m-n)} + \cdots \qquad (99)$$

Next we make provision for the change of scale. The correct substitution must shift the maxima all to the same point (which is best done by shifting them to the origin). This is accomplished by introducing a new variable

$$\delta = n - pm.$$

As pm is the "expectation of n," δ is the deviation from expectation. Next we collapse the scale by setting

$$x = \frac{\delta}{\sqrt{m}} = \frac{n - pm}{\sqrt{m}}.$$

Upon substituting this new variable in (98) and making a few obvious rearrangements, we get

$$P_m(n) = \frac{f(m, n)}{\sqrt{2\pi mp(1-p)}}\left(1+\frac{x}{p\sqrt{m}}\right)^{-pm-x\sqrt{m}-\frac{1}{2}} \times \left(1-\frac{x}{(1-p)\sqrt{m}}\right)^{-(1-p)m+x\sqrt{m}-\frac{1}{2}}. \qquad (100)$$

We now take the two brackets in (100) and treat them by the same process as was used in § 43. If we call their product Z, we have

$$\log Z = -(pm + x\sqrt{m} + \tfrac{1}{2})\log\left(1+\frac{x}{p\sqrt{m}}\right) - [(1-p)m - x\sqrt{m} + \tfrac{1}{2}]\log\left(1-\frac{x}{(1-p)\sqrt{m}}\right).$$

Upon expanding the logarithms in series and collecting like powers of m,

$$\log Z = -\frac{x^2}{2p(1-p)} + \frac{1}{\sqrt{m}}\left(\frac{(p-1)^2-p^2}{p^2(p-1)^2}\,\frac{x^3}{2.3} - \frac{(p-1)+p}{p(p-1)}\,\frac{x}{2.1}\right) - \frac{1}{m}\left(\frac{(p-1)^3-p^3}{p^3(p-1)^3}\,\frac{x^4}{3.4} - \frac{(p-1)^2+p^2}{p^2(p-1)^2}\,\frac{x^2}{2.2}\right) + \dots,$$

whence Z itself is equal to e raised to this power. From this exponential we sort out the term $e^{-x^2/2p(1-p)}$, which is independent of m, and then expand the remainder in a series. The result is

$$Z = e^{\frac{-x^2}{2p(1-p)}}\left[1 + \frac{1}{\sqrt{mp(1-p)}}(2p-1)\left(\frac{1}{2}x - \frac{1}{p(1-p)}\,\frac{x^3}{6}\right) + \left(\frac{1}{\sqrt{mp(1-p)}}\right)^2\left(\frac{8p^2-8p+3}{8}\,x^2 - \frac{7p^2-7p+2}{p(1-p)}\,\frac{x^4}{12} + \frac{(2p-1)^2}{p^2(1-p)^2}\,\frac{x^6}{72}\right) + \left(\frac{1}{\sqrt{mp(1-p)}}\right)^3(2p-1)\left(\frac{8p^2-8p+5}{16}\,x^3 - \frac{94p^2-94p+37}{p(1-p)}\,\frac{x^5}{240} + \frac{10p^2-10p+3}{p^2(1-p)^2}\,\frac{x^7}{144} - \frac{(2p-1)^2}{p^3(1-p)^3}\,\frac{x^9}{1296}\right) + \dots\right].$$

Next, the n in $f(m, n)$ must be replaced by the new variable x, which gives

$$f(m, n) = 1 - \frac{1-p+p^2}{12m\,p(1-p)} + \frac{x(1-2p)}{12m^{3/2}\,p^2(1-p)^2} + \dots.$$

It now becomes obvious that $y = x/\sqrt{p(1-p)}$ would be a better variable to use, so we make this substitution. As $\sqrt{mp(1-p)} = \sigma$ is the standard deviation of the Binomial Law, this new variable satisfies the relation

$$y = \frac{n-pm}{\sqrt{mp\,(1-p)}} = \frac{\delta}{\sigma}, \tag{101}$$

and is therefore the deviation measured in terms of the standard deviation as a unit. We make this substitution in the series for Z and f, and multiply them together, thus getting the result

$$P_m(n) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-y^2}{2}}\left[1 + \frac{1}{\sigma}(2p-1)\left(\frac{y}{2} - \frac{y^3}{6}\right)\right.$$

$$- \frac{1}{\sigma^2}\left(\frac{1-p+p^2}{12} - \frac{3-8p+8p^2}{8}\,y^2 + \frac{2-7p+7p^2}{12}\,y^4\right.$$

$$\left. - \frac{(2p-1)^2}{72}\,y^6\right) + \frac{1}{\sigma^3}(2p-1)\left(\frac{-3+p-p^2}{24}\,y\right.$$

$$+ \frac{47-74p+74p^2}{144}\,y^3 - \frac{37-94p+94p^2}{240}\,y^5$$

$$\left. + \frac{3-10p+10p^2}{144}\,y^7 - \frac{(2p-1)^2}{1296}\,y^9\right) - \ldots\Big]. \qquad (102)$$

The terms which follow are even more complicated.

This distribution function still gives $P_m(n)$, for we have not multiplied it by the Jacobian of the transformation by which we went from n to y. The result of such a change would be to replace $\sqrt{2\pi}\,\sigma$ by $\sqrt{2\pi}$: nothing else would change. This is easily seen from the fact that we have compressed our curve by just the factor $1/\sigma$, and to keep areas unaltered we would need to multiply all ordinates by σ.

Now let us see to what result all this tedious algebra has led us: So long as y^3/σ is small, (102) gives a good approximation to the Binomial Law. Indeed, if y^3 is small enough — which means that we confine our attention to the portion of the curve near the peak — the first term of (102) is good enough. But this term, written for $P_m(y)$ instead of $P_m(n)$, is

$$P_m(y) = \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}}, \qquad (103)$$

which is just the *Normal Law*.

Moreover, when dealing with the special case $p = \frac{1}{2}$, the

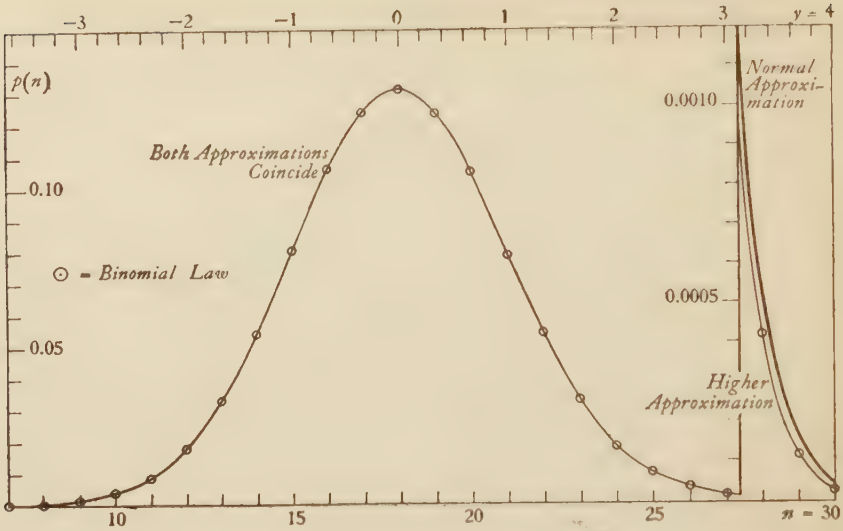


FIG. 24.—APPROXIMATIONS TO THE BINOMIAL LAW.

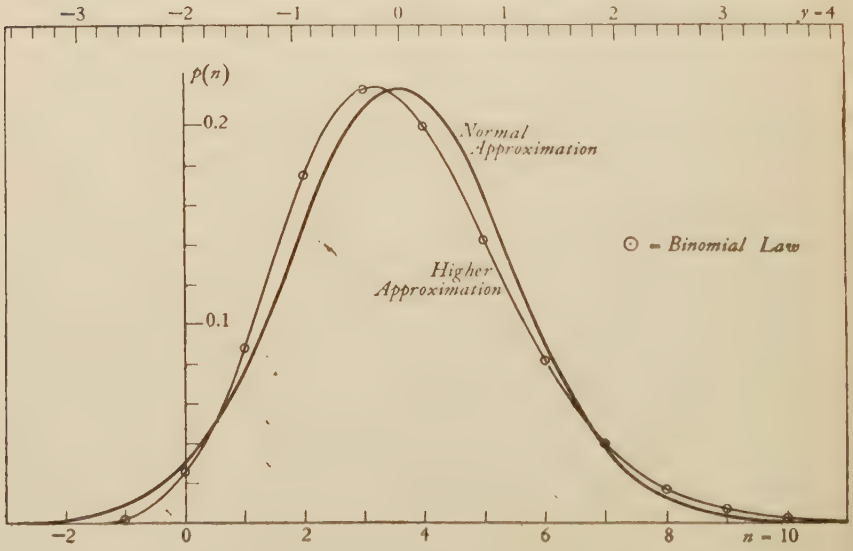


FIG. 25.—APPROXIMATIONS TO THE BINOMIAL LAW.

odd terms in (102) vanish because of the presence of the factor $(2p - 1)$. Hence:

The Normal Law (103) is a fair approximation to the Binomial Law so long as $y^3\ \sigma$ is not too large. In the special case of $p = \frac{1}{2}$ it is somewhat better than otherwise. In the vicinity of the "tails" — that is, when the deviations are large — it is never satisfactory.

A statement of this sort is likely to be rather vague unless it is emphasized by means of some sort of graphical presentation. With this in mind Fig. 24, which corresponds to $m = 36$ and $p = \frac{1}{2}$, and Fig. 25, which corresponds to the same m but to $p = \frac{1}{10}$, are presented. In each case the circles represent the exact values of the Binomial Law; they occur, of course, only at integral values of n. The curves represent our approximations to these values.

Dealing first with the symmetrical case of Fig. 24, we note that the circles and curve coincide absolutely, so far as is possible to judge from the main portion of the drawing. What is more, the rough approximation (103) and the more exact approximation (102) are so nearly coincident that it has been entirely impossible to represent them separately. This would be true right out to $n = 36$ if we were to continue to use the same scale. But if we magnify the ends of the curve, as has been done at the right-hand margin of the drawing, we find that they do indeed separate, and that the higher approximation represents the true values very much better than does the Normal Law (103). For example, at $n = 30$ (the extreme edge of the drawing) the Normal Law is in error by more than 50 per cent, whereas the higher approximation is still indistinguishable from the true value on the scale of the drawing.

Next, turning to Fig. 25, we find that the Normal Law nowhere represents the Binomial with any great degree of exactness, whereas the higher approximation coincides very well throughout the entire range.

§ 83. *Distribution Functions for Discrete Variables; The Poisson Law as a Limiting Case of the Binomial Law*

The second important distribution function in the case of discrete variables is the Poisson Law. It is usually regarded as an approximate form of the Binomial Law when the number of trials m is very large and the probability p very small; and as its derivation from this point of view is the simplest, we shall begin our discussion with it, though we shall shortly see there is another point of view which is of much greater practical importance.

We have found in § 77 that the expectation of n in the case of the Binomial Law is $\epsilon = mp$. Suppose, then, that we replace the p in (23) by ϵ/m, thus causing the formula to take the form

$$P_m(n) = C_n^m \left(\frac{\epsilon}{m}\right)^n \left(1 - \frac{\epsilon}{m}\right)^{m-n}. \tag{104}$$

By writing out the binomial coefficient and rearranging the factors slightly, this expression can be thrown into the alternative form

$$P_m(n) = \left[\left(1 - \frac{1}{m}\right)\left(1 - \frac{2}{m}\right) \cdots \left(1 - \frac{n-1}{m}\right)\right] \times \left[\left(1 - \frac{\epsilon}{m}\right)^{-n}\right]\left[\left(1 - \frac{\epsilon}{m}\right)^{m}\right]\frac{\epsilon^n}{n!}.$$

Now, remembering that we are dealing with a case in which p is supposed to be very small, it is obvious that only those values of n are of consequence which are very small compared to m. Hence every one of the group of terms enclosed in the first set of brackets is of just about unit magnitude. The same is true also of the quantity $1 - \epsilon/m$ which occurs in the second and third brackets, for ϵ/m, or p, is very small. Hence it follows, since there are comparatively few of these terms in the first two brackets, that their product is also not greatly different from unity.

In the case of the third bracket, however, this argument cannot be applied; for the power to which the quantity

$1 - \epsilon/m$ is raised is not a moderate one, but instead is very large. We have seen in § 43, however, that an expression of this form is approximately equal to $e^{-\epsilon}$, so that we are justified in concluding that (104) is equivalent to

$$P_m(n) \doteq \frac{e^{-\epsilon}\epsilon^n}{n!}. \qquad (105)$$

Just how good this approximation is depends upon the values of m, n, and ϵ; of course whatever it is, we could readily improve it to any degree we might desire by the use of processes similar to those in § 82. The result, however, appears to be of little value and probably does not warrant its presentation.

The consequential thing is, that *if p is small enough and m large enough*, the Binomial Law reduces approximately to the form (105), which is exactly the Poisson Law. That these conditions are sometimes satisfied with sufficient approximation to warrant the use of the simpler law we may easily show by citing a particular example which, because of its unusual subject matter, has become classical.

Certain army records, extending over a period of years, give among other things the number of soldiers killed by the kick of horses. The classified results are shown in Table XV. The numbers in the first column represent the number of soldiers killed in this way in one corps during one year, and the second column tells how often this record was repeated during the period covered by the data.

Now there are a large number of days in a year, and the chance of a fatality occurring during any one of these days is pretty small. What is more, each day is a sort of independent "trial"; so there is some reason for expecting the data of Table XV to follow (105) rather closely. To check this supposition there are given, in the third column of the table, the number of times the various records would be expected to have occurred if the distribution accurately obeyed the Poisson Law with an expectation of $\epsilon = 0.61$. We have at present no better means of checking the agreement of the second and third columns than the mere observation that

there does not appear to be any serious disagreement between them. Later on, in Chapter IX, when we have developed a scientific method of measuring this agreement, we shall find that it is very good indeed.

TABLE XV

RECORDS OF SOLDIERS DYING FROM THE KICK OF HORSES

Number of Deaths	Frequency Observed	Frequency Expected
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6
5	0	0.1
6	0	0.0

§ 84. *Definitions of the Phrases "Individually at Random" and "Collectively at Random"*

We can make better use of this illustration, however, than that of merely showing that data sometimes obeys the Poisson Law. We can use it as a guide to the conditions under which the Poisson Law is *exactly*, rather than *approximately*, applicable.

We notice, to begin with, that the times at which the various deaths occurred determine points upon the time axis. If we consider any one death (that is, the death of a particular man), we have no reason to suppose that it occurred at one instant rather than another.[1] That is, it is as likely to fall in one

[1] We must, of course, not be hypercritical about this statement. Certain points on the time axis correspond to periods when the men were asleep, and it is quite unlikely that a death would occur at such an instant. There probably were also different routines established for week days and Sundays, and these might affect the likelihood of the point lying at one place rather than another. Perhaps, however, the fact that what we say about this illustration is *not* true when viewed in too great detail may serve the more clearly to point out the idea we are aiming to convey.

interval of specified length as another; or in the terms of § 55, it is placed on the line " at random."

Then there is the additional fact that the number of deaths occurring during a particular interval is not in any way influenced by what has happened during other intervals. We do not mean by this that there is no sort of connection between them whatever, for obviously there exists the connection implied by the words " the chance of a certain number of deaths in this interval is the same as for any other interval of like length." What we mean can probably best be illustrated by giving an example of the contrary situation. Suppose we were told that in a particular year *just* three men were killed, and that in looking over the records for the first six months we found that exactly two had been killed during that period. Obviously this information does influence our judgment as to what happened in the other half: from the information at our disposal we can conclude that *just one* death occurred in this half. But if we were the statistician who kept the records and at the middle of a year observed a slight excess of deaths for the first half, we would not be able to conclude anything at all about what would happen during the second half. There might be either an excess or a deficit; it is all a matter of chance.[1]

There are many situations about which entirely similar observations might be made. In Chapter X, in dealing with Traffic Problems, we shall have need to return to them again and again. We therefore frame a pair of definitions which shall cover, once for all, the essential ideas with which these observations deal:

---

[1] This statement is also probably not true. An excess observed by a statistician at the middle of the year would probably lead — if it were serious enough — to some sort of "safety campaign" intended to reduce the number of such accidents. Or, viewed from a somewhat closer angle, if the men themselves had immediate knowledge of the occurrence of such deaths, they would probably be led to exert greater care the day following an accident than at other times, and the probability of a death in an interval of a day just following the occurrence of such an accident would be less than normal.

It is just exactly influences of this kind that we wish to rule out in what follows.

A set of points is said to be distributed "individually at random" along a line segment provided each point of the set is placed at random, independently of all the rest.

A set of points is said to be distributed "collectively at random" along a line segment provided the probability of any interval dx containing n points is independent of the number of points in any interval not wholly or partly included in dx.[1]

We have already observed that no set of points can be collectively at random if the total number on the segment is fixed in advance, for under those circumstances an excess in any one interval reduces the chance of an excess, and increases the chance of a deficit, in other similar intervals. This, of course, violates the definition of "collectively at random."

It is equally as easy to set up an illustration in which the points are "collectively" but not "individually" at random. Suppose, for example, that a record is kept of the instants at which people pass a certain point on a bridge. Some of them will come on foot, some will pass in automobiles, and occasionally whole train-loads will pass in subway cars. If we idealize the problem to the extent of ignoring the finite dimensions of conveyances, we may say that, while some persons pass individually, others go by in groups at identical times. The time axis therefore comes to carry a set of points so arranged that the probability of m persons passing during any interval is the same for all intervals.[2] Furthermore, the

[1] This use of the term "collectively at random" must not be confused with the use which has been made of it by Mr. E. C. Molina in connection with variables which can take only discrete values. (See, for example, the *Bell System Technical Journal* for November, 1922.) Mr. Molina's use of the term can be described as follows:

The placing of a point picks out for us some one of the possible values which the variable may take. Similarly, the placing of n points picks out n values. If any such combination of n is equally as likely as any other, the n points are said to be "collectively at random." It is implied, of course, that the same value cannot occur more than once in such a set of n.

[2] The reader will have to approach this illustration in a rather tolerant frame of mind and forget for the moment that there are such things as, on the one hand, traffic police who cause massing of traffic, and, on the other, subway schedules which produce

knowledge of the number of people who have gone by during the last minute (say) has no effect upon the number that are likely to go by during the next minute to come. The points are therefore arranged "collectively at random." But they are not "individually at random," for the point corresponding to a subway passenger is not placed independently of the points corresponding to the other passengers in his train, but is coincident with them.

We thus have illustrations of (*a*), a set of points which is *individually but not collectively* at random, and (*b*), another set of points which is *collectively but not individually* at random. It may be well to round out the ideas by giving a further illustration of a set which is *neither individually nor collectively* at random. Suppose our bridge traffic were regulated in such a way that no two subway trains had a clearance of less than one minute. Then it would certainly be true that the probability of a train-load of passengers passing within the next half minute if a train-load were known to have passed within the last half minute is zero. This violates the definition of "collectively at random." As the illustration still violates the definition of "individually at random," the traffic is not at random in either of these senses.

We add one more illustration of a set of points distributed both individually and collectively at random; this time it is one which is not vitiated by considerations regarding human conduct, as was the one considered at the beginning of the section, but it has its own drawbacks nevertheless.

The emission of β-rays[1] from radioactive substances such

---

a certain regularity of dispersion. The illustration is far from being a perfect one, but it conveys the essential idea better, perhaps, than an artificial system set up for the purpose.

It is a characteristic of abstract logical ideas never to be met in a pure form in actual life.

[1] The term "β-rays" is applied to the electrons spontaneously ejected from the nucleus of an atom. When one is emitted, the substance transmutes itself into a different chemical element, which may have properties quite different from the original one. It may also be radioactive; but as it is no longer the original element, we may truthfully say that an atom emits a β-ray *only once*.

as radium appears to be entirely spontaneous: that is, a nucleus seems impelled to eject an electron, not by what is going on around it, but by some inner "predisposition"—if the word can be allowed—which we do not understand. If we choose any atom and watch it for an interval dt there is a certain probability of it emitting an electron during this interval. As that probability is the same no matter what we may know about *other* atoms, and no matter what interval we choose, the emissions are "individually at random."

If we watch a group of m atoms, the probability of n emissions during dt is absolutely independent of what may have happened during any preceding time,[1] for, as we have said, the emission of β-rays is not affected by events outside the nucleus.

Hence the points marked on the time axis by the successive emissions are both individually and collectively at random.

§ 85. *Second Demonstration of the Poisson Law*

We have seen that it is possible to have sets of points with both of the attributes of randomness defined in § 84, or with neither, or with either one but not the other.[2] We now purpose to show that any set which possesses both these properties is distributed according to Poisson's Law. Specifically, we shall prove the following theorem:

If a set of points is distributed individually and collectively

---

[1] If we chose a given sample of our substance and watched *it*, the atoms of which it is composed would gradually transmute themselves into something else. Hence, as time went on, the number under observation would decrease. We shall find this worth thinking about a little later. For the moment we may suppose that a new atom is somehow fed into the group under observation whenever one leaves it by transmutation.

[2] It is even possible to have a set of points covering a total interval (a, c), but satisfying different conditions within the two parts (a, b) and (b, c). A simple illustration would result (if a, b and c are instants of time) from marking upon the time axis the instants at which β-ray emission took place from a sample of substance, but removing part of the sample at the instant b.

at random in the interval (a, b), *the probability of* n *points lying within any subinterval of length* x *is*

$$P(n, x) = \frac{(kx)^n}{n!} e^{-kx},$$

where kx *is the expected number of points within the subinterval.*

The proof will consist of four parts:

1. That the probability of just one point in an infinitesimal interval dx is an infinitesimal of the same order as dx, while the probability of more than one point in such an interval is an infinitesimal of higher order.

2. That the probability of n points in an interval of length x, when regarded as a function of x, is differentiable.

3. That it has the value stated in the theorem.

4. That kx is the expected number of points in the length x.

1. By definition, the probability that an interval contains n points is the same for all intervals of like length, independently of any information we may have concerning other intervals external to them. This is true of $n = 0$, as of every other value of n.

If we divide the interval x into a number of elements of length dx, x will have no points in it if, and only if, the same is true of every one of its subintervals. Hence, the probability that x is without a point is equal to the probability that every subinterval is without a point.

Let us denote by $P(0, x)$ the probability that there are no points in x; by $P(0, dx)$ the probability that there are no points in dx; and by $P(> 0, dx)$, the probability of one or more points in dx. Then

$$P(0, dx) = 1 - P(> 0, dx)$$

and, since there are x/dx elements in x

$$P(0, x) = [P(0, dx)]^{x/dx} = [1 - P(> 0, dx)]^{x/dx}.$$

This equation only states in symbols what was said in words above: that the interval x is only without points when the same is true of every dx.

Taking logarithms of both sides we obtain

$$\log P(0, x) = \frac{x}{dx} \log [1 - P(> 0, dx)]$$
$$= - x \left[\frac{P(>0, dx)}{dx} + \frac{[P(>0, dx)]^2}{2\, dx} + \dots \right]. \quad (106)$$

We may consider three possibilities:

(*a*) As dx approaches zero $P(> 0, dx)\ dx$ may approach zero. If so, all the remaining terms on the right-hand side of the equation also vanish, which leads to the conclusion that [1]

$$\log P(0, x) = 0,$$

or

$$P(0, x) = 1,$$

independently of the value of x. This would carry with it the conclusion that, no matter how long the interval might be, the probability of its containing a point would be zero. This could only happen provided the points were infinitely far apart. This, of course, is unsatisfactory.

(*b*) Next suppose $P(> 0, dx)\ dx$ approaches infinity. Then *every* term in the right-hand member of (106) is infinite, whence $P(0, x)$ is zero. This means that no matter how small the interval x may be, it is certain to include at least one point. This also is unsatisfactory.

(*c*) The final possibility is that $P(> 0, dx)\ dx$ approaches a limit k which is neither zero nor infinity. In this case $P(> 0, dx)$ approaches $k\, dx$, its higher powers approach higher powers of dx, and all terms except the first vanish in the series (106). This gives

$$\log P(0, x) = - kx,$$

or

$$P(0, x) = e^{-kx}.$$

[1] We know the series to be uniformly convergent for small enough values of $P(> 0, dx)$. Hence we are justified in saying the entire series approaches zero because each term does.

Of the three alternatives, the last is the only one which allows a reasonable value for $P(0, x)$. It must therefore be the correct one. That is, the probability of *one or more points* in the element dx must be an infinitesimal of the first order in dx.

This, however, is not quite what we set out to prove; for it does not necessarily follow that the chance of *more than one* is an infinitesimal of higher order, nor even that $P(1, dx)$ itself is of the first order. The conclusion is merely that *at least* one of the P's vanishes to the first order with dx, and that the rest are infinitesimals of *at least as high order*. If, however, we can show that the probability of two or more points in an interval dx is an infinitesimal of higher order than the probability of one or more, we will have completed the proof that $P(1, dx)$, *and it only*, is of the first order in dx.

To prove this is quite simple. For the event " an interval contains more than one point " is the logical equivalent of the two events " it contains at least one point " and " it contains at least one other." The probabilities of these three events are represented in our scheme of symbolism by the expressions $P(> 1, dx)$, $P(> 0, dx)$ and $P_{>0}(> 1, dx)$, respectively; the last one being the conditional probability of more than one point in an interval in which there is one or more. We have at once, therefore, by the use of (20)

$$P(> 1, dx) = P(> 0, dx)\, P_{>0}(> 1, dx)$$

or

$$P_{>0}(> 1, dx) = \frac{P(> 1, dx)}{P(> 0, dx)}.$$

Now let an interval of length dx be chosen about one of the points. The chance that it contains other points is just $P_{>0}(> 1, dx)$. Let us, then, take the limit of this probability as the length of the interval dx vanishes. The limit is obviously the probability of a second point coinciding with the first, which we have already seen to be zero for points which are placed individually at random. Hence $P_{>0}(> 1, dx)$ must approach zero with dx.

Finally, we notice that

$$P(1, dx) = P(>0, dx) - P(>1, dx)$$
$$= P(>0, dx)[1 - P_{>0}(>1, dx)]$$

whence we conclude that, in the limit,

$$P(1, dx) = P(>0, dx) = k\,dx.$$

This completes our proof of part (1).

2. The second point which we are to prove is that $P(n, x)$ is differentiable with respect to x. To do this we consider two adjacent intervals, one of length x and the other of length dx, the two together constituting an interval of length $x + dx$ when we wish to think of them in combination. Now it is obvious that the compound interval $x + dx$ can only contain n points provided these are distributed between the two component intervals in some one of the combinations

$$n \text{ in } x \text{ and } 0 \text{ in } dx,$$
$$n - 1 \text{ in } x \text{ and } 1 \text{ in } dx,$$
$$n - 2 \text{ in } x \text{ and } 2 \text{ in } dx,$$
$$\cdots \quad \cdots \quad \cdots$$
$$0 \text{ in } x \text{ and } n \text{ in } dx.$$

As the probability of a specified number of points lying in either of these separate intervals is independent of the condition of its neighbor, we conclude that

$$P(n, x + dx) = P(n, x)\,P(0, dx) + P(n - 1, x)\,P(1, dx) + \cdots + P(0, x)\,P(n, dx).$$

However, $P(0, dx)$ must satisfy the relation

$$P(0, dx) = 1 - P(1, dx) - P(2, dx) - \cdots.$$

Substituting this value in the preceding equation and rearranging it slightly we obtain

$$\frac{P(n, x + dx) - P(n, x)}{dx} = [P(n - 1, x) - P(n, x)] \frac{P(1, dx)}{dx} + [P(n - 2, x) - P(n, x)] \frac{P(2, dx)}{dx} + \ldots .$$

We already know that $[P(1, dx)]/dx$ approaches a positive number k as dx approaches zero, and also that $[P(2, dx)]/dx$, $[P(3, dx)]/dx, \ldots$, all approach zero. Hence the entire expression on the right-hand side of the equation approaches a definite limit: the same must therefore be true of the left side also. As this is the necessary and sufficient condition for the differentiability of $P(n, x)$, we have completed our proof.

But we not only know that the derivative exists, we also know its value; for in the limit we have [1]

$$\frac{dP(n)}{dx} = k[P(n - 1) - P(n)]. \tag{107}$$

3. We are now ready for the third point of our proof; that is, for deriving formulæ for $P(n)$. Equation (107) is a linear differential equation; [2] its solution is therefore

$$P(n) = c_n e^{-kx} + k e^{-kx} \int_0^x e^{kx} P(n - 1)\, dx. \tag{108}$$

---

[1] We have no further use for the symbols containing dx. Hence we run no risk of confusion in writing $P(n)$ instead of $P(n, x)$.

[2] Any differential equation of the form

$$\frac{dy}{dx} + f_1(x)\, y = f_2(x)$$

is a "linear differential equation of the first order." Equations of this kind can always be solved, the general formula for the solution being

$$y = e^{-\int_0^x f_1(x)\, dx} \left[C + \int_0^x e^{\int_0^x f_1(x)\, dx} f_2(x)\, dx \right],$$

where C is the constant of integration. How this solution is obtained does not concern us here. Its correctness can be established by substituting it in the original equation.

In our application of it, $y = P(n)$, $f_1(x) = k$ and $f_2(x) = k\, P(n - 1)$. Hence $\int_0^x f_1(x)\, dx = kx$, and (108) follows at once.

Here c_n is the constant of integration, which must be determined in accordance with known conditions of the problem. We have seen, however, that when $n \neq 0$, $P(n)$ vanishes as x becomes zero. But when x is zero, the limits to the sign of integration become equal, and the integral itself vanishes. Therefore all terms in (108) except $c_n e^{-kx}$ are known to vanish, whence it follows that the equation can only be true provided c_n also vanishes.

We can now find the P's one by one. From the first part of our proof we know that

$$P(0) = e^{-kx}.$$

Setting n equal to 1, 2 and 3 in succession in (108) we obtain

$$P(1) = kx\, e^{-kx},$$

$$P(2) = \frac{(kx)^2}{2} e^{-kx},$$

$$P(3) = \frac{(kx)^3}{6} e^{-kx}.$$

This suggests the law

$$P(n) = \frac{(kx)^n}{n!} e^{-kx}. \tag{109}$$

We can establish the truth of this guess by showing that, if it is true for one value, $n - 1$, it must be true for the next higher value, n, also. Upon substituting

$$P(n-1) = \frac{(kx)^{n-1}}{(n-1)!} e^{-kx}$$

in (108) and integrating, the result (109) is obtained. It follows, therefore, that if the law is true for any value of n whatever, it must be true for the next higher value also. It *is* true, however, when n equals 1, 2 or 3. Hence it must be true for every positive integer.

4. The final point in our discussion is to show that kx is

the expectation of the number of points in length x. By definition, this expectation is

$$\epsilon_1(n) = \sum_{n=0}^{\infty} n\, P(n)$$

$$= \sum_{n=0}^{\infty} \frac{(kx)^n}{(n-1)!}\, e^{-kx}$$

$$= kx\, e^{-kx}\left[1 + (kx) + \frac{(kx)^2}{2!} + \ldots\right].$$

The series in brackets is recognized at once as the expansion of e^{kx}. Hence $\epsilon_1(n) = kx$.

This completes our proof, for since kx is the expectation of n, (109) does indeed reduce to the form (105).

§ 86. *Discussion of the Poisson Law; Problems to Which It is an Appropriate Solution*

We have now derived Poisson's Law in two ways, the essential differences between which should not escape notice. In the first place, according to the method used in § 83 the law was merely an approximation which could be safely used provided we knew that the data were distributed in accordance with the Binomial Law, and also that m was very large compared with ϵ. The second method of proof obtains the formula as an *exact solution, not an approximation.* It says nothing whatever about the magnitudes of the numbers x and n, and requires no knowledge as to the number of points within any interval large or small. Instead, it lays down certain general assumptions as to the nature of our knowledge — or, rather, lack of knowledge — concerning the way in which these points follow one after another, asserting: (1) that the probability of one or more points within a definite interval is not influenced by any knowledge we may have concerning the states of other intervals; and (2) that each point in any such interval lies at random, independently of all the rest.

The second method of derivation is frequently a great comfort, as we can see if we consider the incidence of telephone

calls in a telephone exchange. These calls certainly do not fall collectively at random for any great length of time; for the probability of a large number of calls within a minute at three o'clock in the morning is much smaller than in a similar interval at three o'clock in the afternoon. If the incidence of each call were plotted upon a time axis covering an entire day, there would be certain periods in which the points were very dense, and other periods in which they were sparsely scattered. If, however, we choose a period of a quarter of an hour, say, from that part of the day when the traffic appears to be heaviest, or from any other portion except those in which there is a very definite tendency toward change of density, it will be approximately true that any small subinterval has the same likelihood of containing n points as has any other interval of like length. Furthermore, there is very slight dependence among the individual calls. Throughout this quarter of an hour, therefore, the distribution of calls is approximately at random, both individually and collectively. We conclude that Poisson's Formula may be applied *to any time interval whatsoever lying wholly within the quarter of an hour*, or even to the quarter of an hour itself.

This we can infer from our second method of proof. From the first method, on the other hand, we could only infer that the Poisson Law could be applied to an interval which was of sufficiently short duration compared with the quarter of an hour,[1] and we would have no criterion for determining what the words "sufficiently short" mean.

There are many problems of this general type. We have already mentioned the emission of β-rays from a radioactive substance. This is probably the best example in physics, because of the apparently complete independence of the emissions one from another. But there are many others to

[1] If there were known to be exactly m points in the quarter of an hour, and if they were distributed individually at random, the chance of any one of them lying in an interval of length x would be $p = x/15$, if we measure time in minutes. The chance of just n in this interval would then be given exactly by the Binomial Law, to which the Poisson is known to be an approximation *only when p is small.*

which the formula applies in much the same general sense as to the problem of telephone calls. We name only a few typical ones:

The electrons emitted from a hot metal (thermions) or from a photo-sensitive substance under the influence of light (photo-electrons) probably emerge with sufficient independence to make the Poisson Law an excellent approximation when applied to the number appearing within a given interval of time. The number of line surges in a power transmission system because of the throwing of switches undoubtedly falls into the same class, and the number of bursts of static in radio reception probably does. The same is true of demands for service in general, whether upon the cashier of a department store, the stock clerk of a warehouse, or any similar functionary, unless regularity is artificially injected into the system.

Hence it comes about that the Poisson Law is the fundamental basis upon which are solved most of those problems which demand to know the number of persons, or the quantity of apparatus, which will be needed to perform a given service. The number of operators in a telephone exchange, or the number of turnstiles in a subway station, are excellent illustrations. We have not yet arrived at a point where it seems wise to undertake the exact discussion of such problems,[1] but we can with profit consider a very simple one, which, in spite of its simplicity, is so similar to many more practical ones as to aid us in orienting ourselves.

EXAMPLE 47.—*A retail chain store, with limited storage facilities, sells on the average 10 boxes of dog-biscuit per week. The usual practice is to stock up Monday morning. How many packages should be adopted as the standard Monday morning stock, in order not to lose more than one sale out of a hundred?*

If each week is begun with n packages in stock, no sales will be lost unless the demand exceeds n during the week. The chance of a demand for j packages, however, is

$$P(j) = \frac{10^j e^{-10}}{j!},$$

[1] They will be the subject of Chapter X.

for this is exactly the chance of j points (buyers) in unit length (a week) when the expectation for that length is 10.

The expected number of lost sales per week is

$$\epsilon = \sum_{j=n}^{\infty} (j - n) P(j),$$

while the expected number of purchasers is 10. If, then, we were to keep records for a large number of weeks, the number of purchasers would not differ much from $10m$, nor the number of lost sales from ϵm. It follows that, in the long run, the proportion of lost sales would be very close to [1]

$$\frac{\epsilon}{10} = \frac{1}{10} \sum_{j=n}^{\infty} (j - n) P(j).$$

---

[1] This step in our solution presents an excellent chance for error; for there is a treacherous difference between the "expectation of the proportion of sales lost in the long run" and the "expectation of the proportion of sales lost per week."

The best way to see this, perhaps, is by means of a numerical illustration. Suppose that the distribution of customers, instead of following the Poisson Law, were of such a nature that either 8 or 12 appeared each week, one number being equally as likely as the other. Suppose a Monday morning stock of 11 packages were adopted as standard. Then half the weeks would show no lost sales, and half would show one sale lost out of 12. The expectation of the proportion of sales lost per week would therefore be

$$\tfrac{1}{2} \cdot \tfrac{1}{12} + \tfrac{1}{2} \cdot 0 = \tfrac{1}{24}.$$

On the other hand, in the long run, there would be $10m$ prospective customers in m weeks, and $m/2$ lost sales; so that the proportion of sales lost in the long run would be

$$\frac{m}{2} \div 10m = \frac{1}{20}.$$

The two answers are obviously not the same.

The difficulty lies, not in computing the correct answer, for either computation is very simple, but in knowing exactly what it is that we are trying to compute. In the present problem "losing one sale out of a hundred" means that, if we were to keep records over a very long time, the number of lost sales should be about 1 per cent of the number of possible sales, and has nothing whatever to do with the average of the proportions lost during the individual weeks. If, however, we were interested in the latter, we would proceed as follows:

If just j customers appear, the proportion of lost sales is $(j - n)/j$. The chance of this occurring is $P(j)$. Hence the expectation is

$$\epsilon = \sum_{j=n}^{\infty} \frac{j - n}{j} P(j).$$

This formula corresponds to the $\frac{1}{24}$ in our simpler illustration, whereas the one given in the text corresponds to the $\frac{1}{20}$.

To answer the question proposed in Example 47 it is necessary to find the smallest value of n for which this expression is less than 0.01. This we can best do by a process of straightforward computation, but before carrying out the arithmetic it is advisable to throw the formula into a slightly different form. Obviously our formula for $\epsilon/10$ is equivalent to

$$\frac{\epsilon}{10} = \frac{1}{10}\sum_{n}^{\infty} j\,P(j) - \frac{n}{10}\sum_{n}^{\infty} P(j). \tag{110}$$

But

$$\sum_{n}^{\infty} j\,P(j) = \sum_{n}^{\infty} \frac{e^{-10}\,10^{j}}{(j-1)!} = 10\sum_{n}^{\infty} P(j-1).$$

Then, writing $j - 1 = j'$,

$$\sum_{n}^{\infty} j\,P(j) = 10\sum_{n-1}^{\infty} P(j').$$

Substituting this in (110), and noting that it makes no difference whether we call the variable j or j', we obtain

$$\frac{\epsilon}{10} = \sum_{n-1}^{\infty} P(j) - \frac{n}{10}\sum_{n}^{\infty} P(j)$$

$$= P(n-1) + \left(1 - \frac{n}{10}\right)\sum_{n}^{\infty} P(j).$$

The actual computation is shown in Table XVI. The second column contains the values of $P(n-1)$, taken from the table of the Poisson Function given in Appendix VI. The third column is obtained from the second by starting from the bottom and adding the numbers together, each successive partial sum being an entry of the third column. The rest of the computation is obvious.

The least safe stock, under the conditions of the problem, is 16.

TABLE XVI

n	$P(n-1)$	$\sum_{n}^{\infty} P(j)$	$\left(1-\frac{n}{10}\right)\sum_{n}^{\infty} P(j)$	$\frac{\epsilon}{10}$
10	0.12511	0.54207	−0.00000	0.12511
11	0.12511	0.41695	−0.04170	0.08341
12	0.11374	0.30322	−0.06064	0.05310
13	0.09478	0.20844	−0.06253	0.03225
14	0.07291	0.13553	−0.05421	0.01870
15	0.05208	0.08346	−0.04173	0.01035
16	0.03472	0.04874	−0.02927	0.00545
17	0.02170	0.02704	−0.01893	0.00277
18	0.01276	0.01428	−0.01142	0.00134
19	0.00709	0.00719	−0.00647	0.00062
20	0.00373	0.00345	−0.00345	0.00028
21	0.00187	0.00159	−0.00175	0.00012
22	0.00089	0.00070	−0.00084	0.00005
23	0.00040	0.00030	−0.00039	0.00001
24	0.00018	0.00012	−0.00017	0.00001
25	0.00007	0.00005	−0.00007	0.00000

§ 87. *Discussion of the Poisson Law; Variable Traffic Density in a Telephone Exchange*

Both the Poisson Law and the mathematical argument by means of which we derived it in § 85 are more general than the results of that section might lead us to believe. Both, in fact, can be applied to traffic in which there is a very definite trend (that is, to a distribution of points which tend to pack more densely about certain parts of the line than others), provided we know enough about the nature of the trend. We may illustrate this by means of a problem which is not entirely impractical: the incidence of calls in a telephone exchange at a time when the traffic is varying rapidly.

To illustrate exactly what we have in mind, let us think of the highly idealized case of an exchange in which there is absolutely no traffic whatever before nine o'clock in the morning, and in which the traffic density then begins to build up so as to reach a maximum at noon. Suppose we know this to be due, not to a tendency of individual subscribers to place their calls late in the morning, but

to the fact that no subscriber arrives at his place of business before nine o'clock, and many not until considerably later. Suppose, finally, that we have reason to believe that the subscribers arrive at a uniform rate, so that the number at work is a linear function of the time. Under all these highly artificial conditions, we would be justified in the assumption that the chance of a call being made during the short interval of time between t and $t + dt$ is proportional to the number of subscribers then at work, and therefore also to t if nine o'clock is taken as the origin of time.

Now, it is not our intention to deal with so special a case. Instead, we shall be content to assume that the chance of a call coming in during the interval between t and $t + dt$ is represented by some known function of the time, $k(t)\,dt$, which we may suppose to have been arrived at by some such considerations as those given above;[1] and also that the chance of two or more calls in such an interval is an infinitesimal of higher order than $k(t)\,dt$. We shall find that these two assumptions, which are obviously similar to, but very much more general than, our notions of "collectively at random" and "individually at random," respectively, again lead to the Poisson Law.

Let us consider the probability of n calls within an interval of length τ beginning at the instant t. We denote it by the symbol [2] $p(n, \tau, t)$. It is obvious that if we add a very short interval $d\tau$ to the end of τ, we must have the relation

$$\begin{aligned} p(n, \tau + d\tau, t) = {} & p(n, \tau, t)\, p(0, d\tau, t + \tau) \\ & + p(n - 1, \tau, t)\, p(1, d\tau, t + \tau) \\ & + p(n - 2, \tau, t)\, p(2, d\tau, t + \tau) \\ & + \ldots . \end{aligned}$$

This equation assumes nothing except that the probability of a specified number of calls within a particular interval is determined solely by the length of the interval and the time at which it begins, and not by any knowledge which we may have regarding preceding intervals. Furthermore, it is necessarily true that

$$p(0, d\tau, t + \tau) = 1 - p(1, d\tau, t + \tau) - p(2, d\tau, t + \tau) - \ldots .$$

---

[1] This $k(t)\,dt$ will play the same part in our present discussion that the $k\,dx$ played in § 85. There k was the expected number of points per unit length (calls per unit time, or "calling rate"); here $k(t)$ is the "instantaneous calling rate."

[2] This should be read "probability of n [calls] within [an interval of length] τ beginning at [the instant] t."

Substituting this in the preceding equation and making certain minor rearrangements, it becomes

$$\frac{p(n, \tau+d\tau, t)-p(n, \tau, t)}{d\tau}=[p(n-1, \tau, t)-p(n, \tau, t)]\,\frac{p(1, d\tau, t+\tau)}{d\tau}$$
$$+[p(n-2, \tau, t)-p(n, \tau, t)]\,\frac{p(2, d\tau, t+\tau)}{d\tau}$$
$$+\ \ldots\ .$$

If we let $d\tau$ approach zero in this expression, $p(1, d\tau, t+\tau)$ will approach $k(t+\tau)\,d\tau$ by hypothesis, for it is the chance of a new call during the infinitesimal interval $d\tau$ beginning at the time $t+\tau$; while the analogous expressions for two or more calls will be differentials of higher order in $d\tau$. Thus we obtain

$$\frac{d}{d\tau}p(n, \tau, t) = [p(n-1, \tau, t) - p(n, \tau, t)]\,k(t+\tau). \qquad (111)$$

This argument applies to the case of $n = 0$ as well as any other, except that, since negative numbers of calls are absurd, the term $p(n-1, \tau, t)$ drops out. We thus have

$$\frac{d}{d\tau}p(0, \tau, t) = -\,p(0, \tau, t)\,k(t+\tau),$$

whence

$$p(0, \tau, t) = c_0\,e^{-K(\tau, t)},$$

$K(\tau, t)$ being written briefly for

$$K(\tau, t) = \int_0^\tau k(t+\tau)\,d\tau.$$

Returning to the general equation (111), we note that it, like (107), is a linear differential equation, and possesses the solution

$$p(n, \tau, t) = c_n\,e^{-K(\tau, t)} + e^{-K(\tau, t)}\int_0^\tau e^{K(\tau, t)}\,p(n-1, \tau, t)\,k(t+\tau)\,d\tau.$$

The argument by means of which we determined our arbitrary constants in § 85 is of equal force here, and again leads to the conclusion that every c_n is zero except c_0, which is 1. Hence it is only necessary to carry out the successive integrations in order to complete our solution. Taking first the case of $p(1, \tau, t)$ we have

$$p(1, \tau, t) = e^{-K(\tau, t)}\int_0^\tau k(t+\tau)\,d\tau;$$

or, since $k\,d\tau = dK$,

$$p(1, \tau, t) = e^{-K}\int_0^K dK = K\,e^{-K}.$$

Similarly,

$$p(2, \tau, t) = e^{-K}\int_0^K e^K\,p(1, \tau, t)\,dK = \frac{K^2\,e^{-K}}{2!},$$

or in general,

$$p(n, \tau, t) = \frac{K^n\,e^{-K}}{n!}. \qquad (112)$$

Formally, at least, this is the solution of our problem.

In this formula K has a very simple interpretation, similar to the interpretation of kx in (109). There kx meant the "expected number of points in distance x," or, in the terms of our present example, the "expected number of calls in time x." Here also K is the "expected number of calls in the interval τ"; for it is easy to show that

$$\Sigma\, n\,P(n, \tau, t) = K.$$

It is obvious, then, that the Poisson Law has a much wider field of usefulness than would be inferred from the discussion of § 85. We must be careful not to swing to the opposite extreme, however, and conclude that it is *universally* applicable to everything which, in the loose sense in which the phrase is applied in every day speech, occurs "at random." In order to illustrate the type of situation to which it does *not* apply, we consider still another example.

§ 88. *Discussion of the Poisson Law; The General Problem of β-ray Emission*

In a footnote in § 84 we called attention to the fact that as a substance emits β-rays it transmutes itself into a new substance, and thus reduces the amount of the old substance in the sample under observation. This is usually expressed in physics by the assertion that the substance "decays." We now solve the following example:

EXAMPLE 48.—*If the chance of a molecule transmuting itself within a time element dt is $k\,n(t)\,dt$, where $n(t)$ is the number of untransmuted molecules,*[1] *and if there were N molecules present at time $t = 0$, what is the chance of exactly j transmutations between $t = t$ and $t = t + \tau$?*

[1] This law tacitly assumes the products of decay to be non-radioactive.

This problem, like the one in § 87, deals with a set of events which show a definite trend, but there is this vital difference between them: the chance of an event taking place during the interval between t and $t + dt$ was determined solely by t in the preceding case, and was entirely independent of the past history of the system, whereas in the present case it is determined solely by the past history of the system, and (except in so far as the past history is itself different at different instants) is not a function of t. For in the present instance the chance of an emission in the interval in question is proportional to the amount of stuff left, and that amount is completely determined by the number of emissions which have already taken place. The problem is no harder to solve, however, than was Example 47.

We start by investigating the chance of *just n β-rays between* $t = 0$ and $t = t$. We can do this most easily, because we know the number of molecules available for transmutation at both instants, N at $t = 0$, and (since n have been emitted) $N - n$ at $t = t$.

Suppose the interval is lengthened to $t + dt$. Then we have

$$\begin{aligned} p(n, t + dt, 0) = {} & p(n, t, 0)[1 - (N - n)k\,dt] \\ & + p(n - 1, t, 0)(N - n + 1)k\,dt \\ & + \ldots, \end{aligned}$$

the remaining terms being negligible. If we are interested in $n = 0$ this takes the simpler form

$$p(0, t + dt, 0) = p(0, t, 0)(1 - N\,k\,dt).$$

From these relations we easily derive two linear differential equations of which the solutions are [1]

$$p(0, t, 0) = e^{-Nkt},$$

$$p(n, t, 0) = e^{-(N-n)kt} \int_0^t e^{(N-n)kt}\, p(n - 1, t, 0)(N - n + 1)\, k\, dt.$$

From this latter formal solution explicit formulæ can easily be obtained, one by one, for $p(1, t, 0), p(2, t, 0), \ldots$. They are all found to satisfy the law

$$p(n, t, 0) = C_n^N\, e^{-Nkt}\, [e^{kt} - 1]^n. \qquad (113)$$

This, then, is a general formula for an interval beginning at an

[1] The arbitrary constants have been determined exactly as in § 85 and § 87.

instant when the number of molecules is known to be N. What about the interval from t to $t + \tau$ about which our problem asked?

At the time t there must be some number available. The chance that the number is $N - n$ is given exactly by (113). *If* this number is available, the chance of n' being transmuted within the next interval τ is obviously obtained by replacing N and n in (113) by $N - n$ and n', respectively, and t by τ. Hence, making use of the principle of alternative compound probabilities, we have

$$p(n', \tau, t) = \sum_{n=0}^{N-n'} C_n^N e^{-Nkt} (e^{kt} - 1)^n C_{n'}^{N-n} e^{-(N-n)k\tau} (e^{k\tau} - 1)^{n'}.$$

By noting that $C_n^N C_{n'}^{N-n} \equiv C_{n'}^N C_n^{N-n'}$, and making certain other obvious rearrangements, this becomes

$$p(n', \tau, t) = C_{n'}^N e^{-Nk(t+\tau)} (e^{k\tau} - 1)^{n'} \sum_{n=0}^{N-n'} C_n^{N-n'} (e^{k(t+\tau)} - e^{k\tau})^n.$$

The summation is now in the form (15): hence it is equivalent to $(1 - e^{k\tau} + e^{k(t+\tau)})^{N-n'}$. Some further obvious rearrangements then give

$$p(n', \tau, t) = C_{n'}^N \left(\frac{1 - e^{-k\tau}}{e^{kt}}\right)^{n'} \left(1 - \frac{1 - e^{-k\tau}}{e^{kt}}\right)^{N-n'}. \qquad (114)$$

This is the general solution of the problem. *It is not in the form of Poisson's Law, and cannot be reduced to that form, the reason being that the chance of a β-particle being emitted in any interval is very definitely dependent upon the occurrences in other intervals.*

It is only when N is very large and k (the "rate of decay") is very small (so that Nk is of moderate size), that it reduces to Poisson's form. Under these circumstances the number left after any physically consequential time t is not very greatly different from N, so that the chance of an emission between t and $t + dt$ is virtually independent of t.

§ 89. *An Approximation to the Poisson Law*

It can be shown that under certain circumstances the Normal Law is an acceptable approximation to the Poisson Law. We begin with (105), but omit the subscript m from our symbol, since we are no longer thinking of it as an approximate form of the Binomial Law. We thus get

$$P(n) = \frac{\epsilon^n e^{-\epsilon}}{n!}. \qquad (115)$$

As usual, we denote the deviation from expectation by δ and the standard deviation by σ. A simple algebraic computation then shows that $\sigma^2 = \epsilon$.

From this point on, our procedure is exactly parallel to that carried out in § 82. We first replace n by the new variable $y = \delta/\sigma = (n - \sigma^2)/\sigma$, which measures the deviation in terms of the standard deviation as a unit. We thus get

$$P(n) = \frac{\sigma^{2\sigma^2 + 2\sigma y}}{(\sigma^2 + \sigma y)!} e^{-\sigma^2}.$$

We next replace the factorial by its Stirling approximation and expand the various terms in series. The final result, after much tedious algebra, is

$$P(y) = \frac{e^{\frac{-y^2}{2}}}{\sqrt{2\pi}}\left[1 - \frac{1}{\sigma}\left(\frac{y}{2} - \frac{y^3}{6}\right) - \frac{1}{\sigma^2}\left(\frac{1}{12} - \frac{3y^2}{8} + \frac{y^4}{6} - \frac{y^6}{72}\right)\right.$$
$$+ \frac{1}{\sigma^3}\left(\frac{y}{8} - \frac{47y^3}{144} + \frac{37y^5}{240} - \frac{y^7}{48} + \frac{y^9}{1296}\right)$$
$$\left. + \frac{1}{\sigma^4}\left(\frac{1}{288} - \frac{5y^2}{32} + \frac{347y^4}{1152} - \frac{617y^6}{4320} + \frac{23y^8}{960} - \frac{y^{10}}{648} + \frac{y^{12}}{31104}\right) + \ldots\right]. \tag{116}$$

The first term of this series is exactly the same as the first term of the series for the Binomial approximation (102). That is, *the Normal Law (103) is a satisfactory approximation to the Poisson Law when y^3/σ is not too large.*

To see just how satisfactory the approximation is, we refer to Figs. 26 and 27, in which the values of the Poisson Law are represented by circles, while the Normal Law and the more complete approximation (116) are shown as continuous curves. In Fig. 26 the value of ϵ is 10, and in Fig. 27 it is 100. The discrepancies between the true values and those given by the Normal Law are considerable in every case, though much more noticeably so for the smaller value ϵ than for the larger one; but near the center of the range, where the

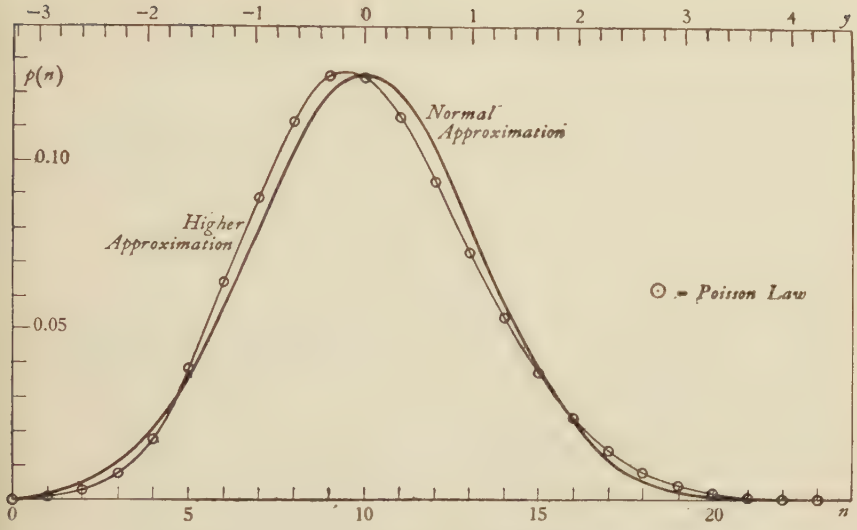


FIG. 26.—APPROXIMATIONS TO THE POISSON LAW.

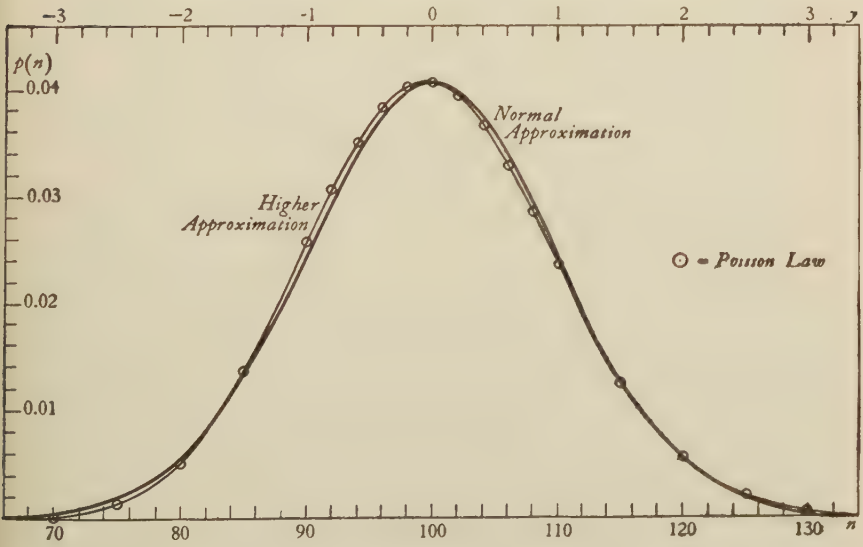


FIG. 27.—APPROXIMATIONS TO THE POISSON LAW.

probabilities are high, the percentage error would not be of serious consequence for many purposes. Near the tails, however, the percentage error is very large. Therefore the approximation should never be used in those regions. The complete approximation (116), on the other hand, agrees very well with the true values throughout the entire range, even in the case of the smaller value of ϵ.

PROBLEMS

1. Suppose that telephone calls, each of length T, occur individually and collectively at random, the calling rate being n per unit time. What is the probability that, at the instant t, there are *exactly* j in progress?

2. What is the probability that *more than* j are in progress?

3. What is the chance of *two or more* calls in an interval of length dt? If every call is followed by a "danger interval" — that is, an interval of length dt within which another call, if it arrives, will interfere with the first — what is the probability that a call will be interfered with during its danger interval? Why are these answers not equal?

4. At the time 0 an observer begins to note the arrival of calls. What is the probability that the *first* call arrives between t and $t + dt$?

5. What is the probability that the interval between a call and its next succeeding call lies between t and $t + dt$?

6. What is the expected time of waiting in Problem 4? The expected time interval in Problem 5?

7. With respect to Problem 6 the following argument can be made: The time $t = 0$, at which the observer enters, must lie in an interval between calls. It is just as likely to lie near the beginning as the end of the interval. That is, its average position is the middle of the interval. Hence the average waiting time of the observer will be half the average interval between calls.

The correct answers to Problem 6 do not satisfy this condition. Explain the paradox.

8. Suppose an exchange is suddenly "cut into service" at the height of busy hour traffic. Call the instant $t = 0$. For negative values of t the calling rate is zero. Hence there are no calls in any interval. For $t > 0$ the calling rate is n per unit time: hence Poisson's

Law applies to any interval lying wholly in this time. But if an interval begins at a time $t < 0$ and ends at $t + \tau > 0$, the calling rate is not constant. What is the probability of just j calls in such an interval?

9. In Problem 8, assume $n = 3$, $\tau = 1$, and draw curves for $P_j(\tau, t)$ covering values of t from -2 to $+1$, for each of the following values of j: 0, 1, 2, 3. Discuss the characteristic features—particularly maxima—of these curves.

10. In Problem 5, what is the most probable length of interval? The root mean square length? The expected length?

11. In the decay exercise of § 88, what amount of substance can be expected to remain after a time t?

12. What *proportion* of the amount present at time t decays during the next second? Note that t does not appear in your answer.

13. What is the expected time of emission of the first β-particle?

§ 90. *The Normal Law*

So far in this chapter we have been dealing with two distribution laws which appear as the consequences of sets of assumptions to which the circumstances of engineering problems frequently lead us; or at least from which the circumstances of engineering frequently differ in minor respects, the importance of which we can to some extent estimate. We now come to the consideration of a totally different class of distribution functions — no less important, but justified solely by their proven utility and not in any way by theoretical appropriateness.

First in importance among these we must list the Normal Law (103). It may seem strange to refer to it in that fashion after we have twice assigned it more specific significance — once as an exact law under assumptions that at least remotely resemble the facts of molecular velocities in a gas, and then again as an approximate form of either the Binomial or the Poisson Law. But the use of the Normal Law is not confined to such cases. In addition, there are many problems which involve unknown distribution functions — virtually every case of the deviation of manufactured parts from the ideal qualities

which they are designed to have is of this type — and the best that we can usually do under such circumstances is to choose some function which appears to possess the major characteristics that experience has taught us to expect in such distributions, and then to arbitrarily fit our data to it. It is in such cases that the Normal Law plays the part of a purely empirical law.

Attempts have frequently been made to justify its use in this field, but they have rather signally failed to do so on two scores: in the first place, experience has shown that most such distributions do not conform to even the major characteristics of the law; in the second, the assumptions upon which the justifications are based are of such an intangible character that it is impossible to predict in advance whether or not a given scientific situation is one of the appropriate few. Derivations based upon postulates which have no concrete physical interpretation are of little use to the scientist. We shall therefore not concern ourselves with them.[1]

Instead, let us write down the principal characteristics of the law; for if we cannot determine *a priori* whether a certain set of data ought to conform to it, we can at least check up the data, once it has been obtained, to find whether it is of a radically different type. The principal characteristics are:

1. The Normal Law is symmetrical. Negative and positive deviations of like magnitude are equally likely to occur.

2. The Normal Law assigns a finite probability to every finite deviation. There are no excluded cases.

3. In the case of the Normal Law there is just one most probable result, and this is identical with the first expectation of the variable.

---

[1] I do not mean to infer that the study of such proofs is necessarily a waste of time: I only mean that (unless it contributes to the advancement of the logic or mathematics *per se*, with which we are not concerned) it belongs rather to the category of scientific recreations than to that of purposeful work. The student who is interested in following up the subject will find three excellent, and entirely different, derivations in the References for Outside Reading at the end of the chapter.

Now if a set of data is to be distributed in accordance with the Normal Law it must possess all three of these characteristics, and it requires only a moment's consideration to see that the possession of all of them is by no means general. Take, for example, the case of the height of men. It is absurd to speak of a man of negative height: such a thing is not merely very infrequent; it simply cannot occur. Yet if height were distributed normally, the second property of the Law would assign a finite probability to this absurdity. Or, as a further example, we may think of the lengths of telephone conversations. Here again negative values are meaningless; but more than that, actual experience has shown that the conversation of most frequent occurrence is very short indeed, certainly much shorter than the average. The distribution function which these call-lengths obey must therefore possess neither of the three properties.[1]

It is an unescapable consequence of these facts that, if statistical investigations are to have any degree of generality at all, they must deal with other distribution laws as well, and in the few remaining sections of this chapter we shall present briefly some of the more successful attempts that have been made to broaden out the material with which the statistician may work. But before passing on to this subject, we must pause for a moment to consider wherein the importance of the Normal Law really lies. We have said that as an exact law its demonstrations have been of a sadly impractical character, and that as an empirical law its own peculiarities are of so special a type that it is seldom obeyed; and it might appear that this left no field of usefulness for it whatever. To form this conclusion, however, would be grossly unfair, for it possesses very considerable virtues nevertheless.

In the first place, it depends upon a single variable y, and is therefore very easy to tabulate. In contrast, the Binomial

[1] In fact, in this particular case a much better law appears to be of the form $P(T) = a e^{-bT}$ for positive values of T, and $P(T) = 0$ when T is negative. See in this connection a recent paper by Mr. E. C. Molina in the *Bell System Technical Journal* for July, 1927.

Law depends upon two variables m and n, the Poisson Law upon ϵ and n, and the law (25) upon m, n, p and q. And this is no mean advantage, for it means that the Normal Law is the simplest of all laws to handle in those cases to which it does apply. In fact, its superiority in this respect is great enough to warrant its use in many places where a better approximation to the true distribution could undoubtedly be obtained, but where a rough estimate easily gotten is preferable to a more exact one which demands far more labor.

In the second place, as we shall see in our discussion of the Gram-Charlier series, it is possible to express many distributions which do not possess the three properties listed above in terms of the Normal Law and its successive derivatives; and since the derivatives have also been tabulated, this often affords a feasible means of securing greater exactitude.

In the third place, we have shown that both the Binomial and the Poisson Laws, under suitable conditions, approach the Normal Law as a limit. In § 101 we shall develop an analogous theorem of much greater generality; and it has, in fact, been shown that a very wide variety of distributions have the Normal Law as a sort of common limit to which they all tend. This property is probably the most important one of all, for though it may be of little use to us in actual computation, it seems to say that the Normal Law somehow epitomizes that element of accidental distributions which is common to all sorts of circumstances. It is, in a sense, the Center of Perspective of probability.

§ 91. *Empirical Families of Curves; Pearson's Curves*

Among the families of curves which have been set up for the purpose of enabling the statistician to deal with a wide variety of data, the most empirical of all is that developed by Karl Pearson, for the foundation underlying it is only suggestive at best. This foundation consists of the observation that, in a certain approximate sense which will shortly appear, the Normal Law, the Binomial Law, the Poisson Law and the law

of repeated dependent trials given in (25) all satisfy the differential equation

$$\frac{1}{P}\frac{dP}{dx} = \frac{a + x}{b + cx + dx^2} \tag{117}$$

for some set of values of the constants a, b, c, d. Certain solutions of this equation are then sorted out, largely because of their algebraic simplicity, to form the desired family of distribution types.

Take, for example, equation (23), which was represented graphically in Fig. 7. If we join the tops of consecutive ordinates by straight lines, as shown in Fig. 28, we produce a polygon, or sort of "broken curve." Of course, the origin of the Binomial Law is such that the only values of n to which any significance can be attached are those which correspond to the "corners" of this "curve"; but we are going to ignore this fact for a moment, and think only of its appearance as a geometrical diagram. It is obvious that if we were to draw a smooth curve, not through the corners themselves, but through the mid-points of the connecting lines instead, and if we required this curve to be tangent to each connecting line at its mid-point, we would get something which possessed the principal characteristics of the polygon itself. We can actually carry out this process algebraically by noting that the abscissa and ordinate of the mid-point are, respectively,

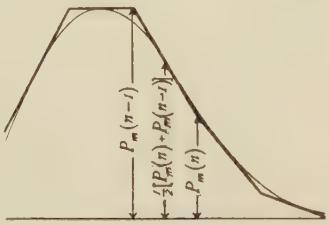


Fig. 28.

$$x = n - \tfrac{1}{2}$$

and

$$P = \tfrac{1}{2}[P_m(n) + P_m(n-1)],$$

and that the slope of the connecting line is

$$P_m(n) - P_m(n-1).$$

From these values we easily find that at each mid-point the smooth curve of which we have spoken would satisfy the differential equation

$$\frac{1}{P}\frac{dP}{dx} = 2\,\frac{P_m(n) - P_m(n-1)}{P_m(n) + P_m(n-1)},$$

which, upon replacing $P_m(n)$ and $P_m(n-1)$ by their algebraic expressions as given by (23), and changing the variable from n to x, reduces at once to

$$\frac{1}{P}\frac{dP}{dx} = \frac{(\frac{1}{2} - mp - p) + x}{\left(-\frac{1}{4} - \frac{mp}{2}\right) + (p - \frac{1}{2})x + (0)x^2}.$$

This equation is of exactly the form (117), the quantities which correspond to the constants a, b, c and d being enclosed in parentheses.

It is in this approximate sense that the Binomial Law may be said to obey the differential equation (117).

It is interesting to note in passing that if $p = \frac{1}{2}$, which makes the Binomial Law symmetrical, and if we write $y = (x - pm)/\sqrt{(m+1)p(1-p)}$ which is very much like the substitution used in (93), our equation becomes

$$\frac{1}{P}\frac{dP}{dy} = -y,$$

of which the solution is just the Normal Law [1]

$$y = \frac{1}{\sqrt{2\pi}}\,e^{\frac{-y^2}{2}}.$$

In the case of the law (25) also we are led to the differential equation (117), provided we assume the supply of balls of each color (m and n) and the total number drawn ($p + q$, which we call r in what follows) to be specified, the variable

[1] The constant of integration must be put equal to $1/\sqrt{2\pi}$, otherwise the area under the curve would differ from unity.

being the number of one color obtained (that is, p). In this case, however, none of the constants is zero. They are, in fact

$$a = -\frac{2mr + m + 2r - n}{2(2 + m + n)},$$

$$b = -\frac{2mr + m + n + 1}{4(2 + m + n)},$$

$$c = \frac{2r + m - n}{2(2 + m + n)},$$

$$d = -\frac{1}{2 + m + n}.$$

The demonstration of the fact that the Poisson Law also, in the same approximate sense, obeys the differential equation (117) is left as an exercise for the student.

In a vague sort of way, therefore, (117) is characteristic of all four of these important types of distribution, and it is natural to expect that an empirical family whose members are bound together by this same law may possess advantages over one with even less by way of logical foundation. It was from considerations such as these that Pearson was led to examine all the solutions of (117) with a view to finding what sort of data they are capable of representing.

The solutions in question, however, take different forms according as the quantity $c^2 - 4bd$ is positive, negative or zero. They are: if $c^2 - 4bd$ is positive,

I–II. $$P = k(x - \alpha_1)^{m_1}(x - \alpha_2)^{m_2}; \tag{118}$$

if $c^2 - 4bd$ is negative,

IV. $$P = k[(x - \alpha)^2 + \beta^2]^m e^{-\gamma \tan^{-1}\frac{x-\alpha}{\beta}}; \tag{119}$$

and if $c^2 - 4bd$ is zero,

$$P = k(x - \alpha)^m e^{-\frac{a+\alpha}{x-\alpha}}. \tag{120}$$

In addition there are degenerate forms: when $d = 0$,

III. $$P = k(x - \alpha)^m e^{-\beta x}; \tag{121}$$

and when both c and d are zero,

$$P = k e^{-\beta(x-\alpha)^2}. \tag{122}$$

It need hardly be said that all the constants which appear in these equations, with the exception of the constant of integration k, are written in place of combinations of a, b, c and d, and are therefore perfectly arbitrary so long as a, b, c and d are. As for the k in any equation, it must have such a value that the area under the curve defined by that equation is unity.

Pearson chose the first of these as his "Type II" curve in the special case when $m_1 = m_2$; when $m_1 \neq m_2$ he called it "Type I"; the next he called "Type IV" in the special case when $m = -1$; and as "Type III" he used equation (121). He also listed several other cases, but these will be quite enough for our consideration.

Let us now see what are the characteristics of the various Types:

Taking first Type II, it is obvious that P either vanishes or becomes infinite at $x = \alpha_1$ and $x = \alpha_2$: which of the two happens depends upon the sign of the exponent, but whichever happens at α_1 also happens at α_2, for it must be remembered that in the case of Type II the exponents are equal. In no case, however, may the exponent be less than -1, for if it were, the area under the curve would not be finite.[1]

[1] It will be remembered that integrals in which the integrand becomes infinite — they are commonly called "improper integrals" — are defined as limits of integrals which do not contain the troublesome point. Thus

$$y = \int_0^a \frac{dx}{x^m}$$

is defined to mean

$$y = \lim_{b \doteq 0} \int_b^a \frac{dx}{x^m}.$$

But

$$\int \frac{dx}{x^m} = \frac{1}{1-m} x^{1-m};$$

Hence

$$y = \lim_{b \doteq 0} \frac{a^{1-m} - b^{1-m}}{1-m}$$

which is either $\frac{a^{1-m}}{1-m}$ or ∞, according as $m < 1$ or $m > 1$.

The same argument applies to the attempt to integrate (118): if either m_1 or m_2 is less than -1, the area under the curve is infinite, no matter what value may be given to k.

The same remarks apply to Type I also, except that it is no longer necessary that the curve behave in the same way at both end points. Instead it is quite possible for one exponent to be positive and the other negative, in which case the function vanishes at one end and becomes infinite at the other. Moreover, the curves, whether of Type I or Type II, have quite different properties according as the m's lie in the intervals

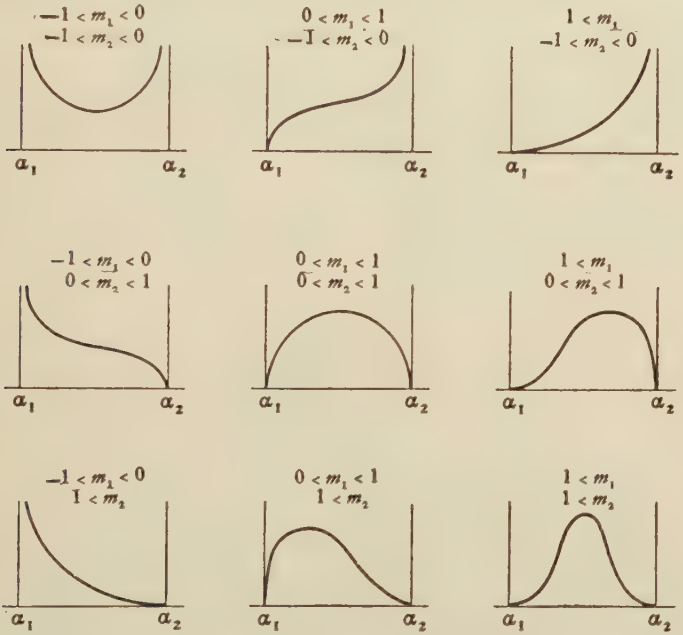


Fig. 29.—Typical Forms of Pearson Curves. Types I and II.

$-1 < m < 0$; $0 < m < 1$; $1 < m$. In the first case, the curve becomes infinite at the end point; in the second it becomes zero, diving into the axis abruptly; in the last case it becomes zero, but meets the axis smoothly. As either of these conditions at one end can be combined with either at the other, nine typical results can be obtained, as shown in Fig. 29.

Of course, the Type II curves are the special *symmetrical* cases of Type I.

It is obvious that these curves cover a pretty wide range of

"shapes." They have, however, the common property of excluding values outside the range (α_1, α_2) completely. We still need curves which are limited at one end only, and others which are not limited at all.

Type III supplies the first sort.[1] Again we must distinguish between values of m in the three ranges $(-1, 0)$, $(0, 1)$, $(1, \infty)$; and also between positive and negative values of β. The various forms which this type may assume are shown in Fig. 30.

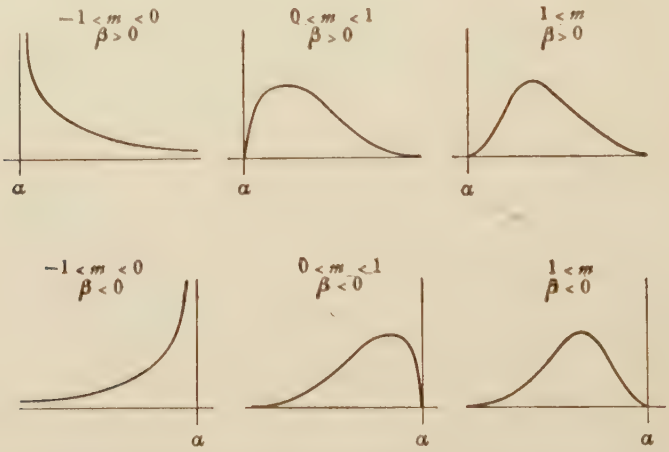


FIG. 30.—TYPICAL FORMS OF PEARSON CURVES. TYPE III.

Finally, the curves resulting from equation (119) extend to infinity in both directions. They can only define a finite area, however, provided $m < -\frac{1}{2}$. Pearson chooses $m = -1$, and writes the equation [2]

$$P = \frac{k\, e^{\gamma \tan^{-1}\left(\frac{x}{\beta}\right)}}{1 + \frac{x^2}{\beta^2}};$$

[1] Pearson sets $m = -\beta\alpha$. This looks like a restriction on the value of m, but really is not; for any possible curve (121) can be made to coincide with one for which $m = -\beta\alpha$ by simple translation.

[2] Again, (119) degenerates by simple translation into an equation without the α.

this value of m being particularly desirable because

$$\frac{d}{dx}\tan^{-1}\frac{x}{\beta} = \frac{1}{\beta}\,\frac{1}{1+\frac{x^2}{\beta^2}}.$$

The two shapes which characterize this type are gotten when γ and β have the same sign ($\gamma\beta > 0$), and when they have not ($\gamma\beta < 0$). They are shown in Fig. 31. Neither of the two is symmetrical; but there is a degenerate special case when $\gamma = 0$ which is symmetrical. It, too, is shown in Fig. 31, in comparison with the Normal Curve (which is dotted).

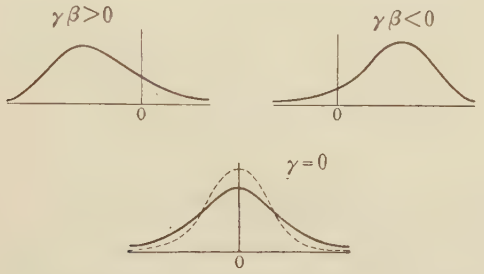


Fig. 31.—Typical Forms of Pearson Curves. Type IV.

Obviously, the Pearson Type IV gives a much higher chance of large deviations than does the Normal Law.

That the Pearson Curves cover a wide range of possibilities is obvious at sight, and experience has shown that they can be made to fit a large proportion of statistical data surprisingly well. How they are used will be discussed in Chapter IX. We now turn our attention to other families of curves.

§ 92. *Empirical Families of Curves; The Gram-Charlier Series*

We now come to the discussion of a totally different method of obtaining a distribution function to represent a set of experimental data — one which starts from a basis which is much sounder logically than is that underlying the Pearson Curves.

We start with the Normal Law (103), which for the moment we shall write in the form

$$\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

in order to conform to the notation which has come to be more or less standard in dealing with this subject. It is a simple matter to show that its successive derivatives are

$$\phi'(y) = \frac{-1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} y,$$

$$\phi''(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} (y^2 - 1),$$

$$\phi'''(y) = \frac{-1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} (y^3 - 3y),$$

or in general,

$$\phi^i(y) = \frac{(-1)^i}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} H_i(y),$$

where $H_i(y)$ is written briefly for the "Hermite Polynomial"

$$H_i(y) = y^i - 1\, C_2^i\, y^{i-2} + 1.3\, C_4^i\, y^{i-4} - 1\cdot 3\cdot 5\, C_6^i\, y^{i-6} + \ldots.$$

These H's and ϕ's possess the remarkable property that the integral of the product $H_i(y)\,\phi^j(y)$, taken from $-\infty$ to $+\infty$ is zero, no matter what the values of i and j may be, so long as they are not equal. This property, which is known in mathematics as "biorthogonality," is always a very valuable one; for it permits us to make use of a very simple method of expanding an arbitrary function, $f(y)$, into a series of the form

$$f(y) = c_0\,\phi(y) + c_1\,\phi'(y) + c_2\,\phi''(y) + \ldots. \qquad (123)$$

To show this, suppose we multiply both sides of this equation by $H_i(y)$ and then integrate the result term by term between the limits $-\infty$ and $+\infty$. Because of the fact

that the functions are biorthogonal, every term on the right-hand side of the equation will vanish except the term for which the two indices are equal, thus giving us the relation

$$\int_{-\infty}^{\infty} H_i(y)\, f(y)\, dy = c_i \int_{-\infty}^{\infty} H_i(y)\, \phi^i(y)\, dy,$$

or

$$c_i = \frac{\int_{-\infty}^{\infty} H_i(y)\, f(y)\, dy}{\int_{-\infty}^{\infty} H_i(y)\, \phi^i(y)\, dy}.$$

Now, by straightforward integration we find that the denominator of this expression is equal to $(-1)^i i!$; wherefore

$$c_i = \frac{(-1)^i}{i!} \int_{-\infty}^{\infty} H_i(y)\, f(y)\, dy.$$

In other words, *the coefficients in the series which represents* $f(y)$ *are simply the product of known numerical factors by the integrals from minus infinity to plus infinity of the products of the Hermite polynomials into the function which is to be expanded.*

Of course all this is true only provided it is *possible* to expand $f(y)$ into such a series; and of course it is only useful provided we are able to carry out the integrations. So it will be wise for us to give our attention for a moment to each of these questions.

As for the first, it can be shown that any function $f(y)$ which, together with its first two derivatives, is continuous, and all the derivatives of which vanish at infinity, is capable of representation by means of such a series. These conditions would seem to be broad enough to cover most statistical studies.

As for the second, one condition is very obvious: we cannot evaluate the integrals in question unless we know what the function $f(y)$ is, *and in the case of statistical studies we almost*

never do. What we usually know is, that certain observations have given us certain results. They admittedly do not represent the function exactly; but they are the best we have, and the problem before us is that of making the very best possible use of them. In the attempt to do this, we shall be much better pleased with a result that does not obviously disagree with the data, than we will with another result that still is grossly in error even after we have done our best with it, no matter how sound the latter may be theoretically.

We can perhaps illustrate this idea a bit better by reference to a more familiar type of expansion. It is well known that the function e^{-x^2} can be represented for every value of x by means of a Taylor's Series. If, however, we possessed data which obeyed this law, though we did not know it did, and if our data were only extensive enough to permit the determination of three coefficients in our Taylor's Series, we would be hard put to it to find a series which possessed even the major characteristics of the function in question. Certainly no three-term polynomial would do. But we might get something of practical utility from the use of the function $a \cot^{-1}(b + c x^2)$ which, theoretically, is not the right thing to begin with at all.

The same general situation exists with reference to the Gram–Charlier Series. So far as their application to statistical data is concerned, fine-spun theories regarding convergence and the like are likely to be a work of supererogation; for it is usually possible to determine no more than three or four coefficients at the most, and the practical question is simply how far the simple expressions thus derived are capable of representing our data.

But if the series have no substantial theoretical foundation, so far as their use in representing empirical data is concerned, they do have very substantial merits of another sort. For one thing, it has actually been found by experience that even the few terms with which we may conveniently deal are capable of representing many sorts of data in a quite satisfactory fashion; for another, they are easy to compute because,

as we have already said in our discussion of the Normal Law, extensive tables of the functions $\phi^i(y)$ are available to assist us.[1] The Pearson types, on the contrary, are sometimes quite unwieldy.

How the coefficients are to be determined when the series are used in this empirical fashion will appear in Chapter IX. For the moment we return to the realm in which the theoretical foundation of the series does apply — the realm where the true law is known in advance.

§ 93. *Gram-Charlier Approximations to the Binomial and Poisson Laws*

In § 82, in an attempt to derive a formula which would approximate the Binomial Law, we arrived at the formidable series (102), the computation of which in any individual case is likely to be tedious. But by comparing the coefficients in this series with the known expressions for the Hermite polynomials[2] it is a simple

---

[1] We should also mention the fact that the Gram-Charlier Series has been deduced from certain assumptions regarding "elementary errors" analogous to those which lead to the Normal Law, and therefore possesses whatever merit such an argument implies.

I do not believe it establishes the logical preeminence of the Gram Series as a medium for the representation of distributions of data; but, as Dr. W. A. Shewhart has pointed out, if the fluctuations in a variable can be represented accurately by one or two terms of a Gram Series, the causes underlying those fluctuations must be *of the same effect* as if they were actually "elementary errors." When this situation exists, no amount of data is likely to teach us much about the individual causes of variation; when it does *not* we may hope to learn something about those causes, and, if the variation be an undesirable one, to eliminate them.

[2] We have already given in § 92 a general formula for the Hermite polynomials. We add a list of the first thirteen:

$$H_0 = 1, \qquad H_5 = y^5 - 10y^3 + 15y,$$
$$H_1 = y, \qquad H_6 = y^6 - 15y^4 + 45y^2 - 15,$$
$$H_2 = y^2 - 1, \qquad H_7 = y^7 - 21y^5 + 105y^3 - 105y,$$
$$H_3 = y^3 - 3y, \qquad H_8 = y^8 - 28y^6 + 210y^4 - 420y^2 + 105,$$
$$H_4 = y^4 - 6y^2 + 3, \qquad H_9 = y^9 - 36y^7 + 378y^5 - 1260y^3 + 945y,$$
$$H_{10} = y^{10} - 45y^8 + 630y^6 - 3150y^4 + 4725y^2 - 945,$$
$$H_{11} = y^{11} - 55y^9 + 990y^7 - 6930y^5 + 17325y^3 - 10395y,$$
$$H_{12} = y^{12} - 66y^{10} + 1485y^8 - 13860y^6 + 51975y^4 - 62370y^2 + 10395.$$

matter to throw it into the form

$$P_m(y) = \phi + \frac{2p-1}{6\sigma}\phi''' + \frac{1}{\sigma^2}\left(\frac{6p^2-6p+1}{24}\phi^{iv} + \frac{4p^2-4p+1}{72}\phi^{vi}\right)$$
$$+ \frac{2p-1}{\sigma^3}\left(\frac{12p^2-12p+1}{120}\phi^{v} + \frac{6p^2-6p+1}{144}\phi^{vii}\right.$$
$$\left. + \frac{4p^2-4p+1}{1296}\phi^{ix}\right) + \ldots . \qquad (124)$$

This series, which is of the Gram-Charlier type, is much easier to evaluate than (102) because of the aid that may be gotten from tables such as those in Appendix IV. It is particularly useful when we desire to know the probability of n lying between two preassigned limits n_1 and n_2. To see why, we refer back again to Fig. 25, in which the circles represent the actual values of $P_m(n)$, while the smooth curve which passes through them gives the approximation (102), or what amounts to the same thing, the Gram-Charlier Series (124). We reproduce part of this figure as Fig. 32, except that the true values of $P_m(n)$, instead of being plotted as points, are represented by rectangles as in Fig. 8.

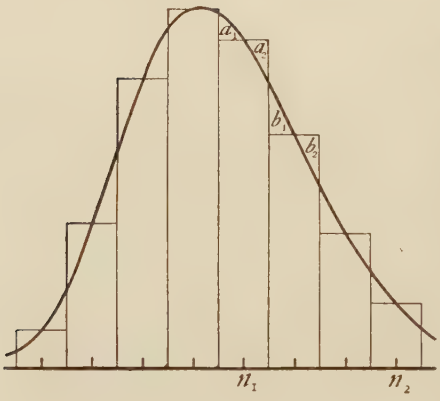


FIG. 32.

It is obvious that the area under the smooth curve between the ordinates at n_1 and n_2 is smaller than the area of the corresponding set of rectangles,[1] for the rectangles extend a half unit beyond these ordinates in each direction. On the other hand, the area under the smooth curve between the limits $n_1 - \frac{1}{2}$ and $n_2 + \frac{1}{2}$ is an excellent approximation to the area of the rectangles, since the little triangular areas a_1, a_2; b_1, b_2; . . . , pretty nearly compensate for one another.

We can, therefore, find a pretty good approximation to the

[1] If the meaning of this is not clear, the reader is refered to the last three paragraphs of § 38.

probability of n lying between n_1 and n_2 by taking the integral of (124), using as limits of integration, not the values y_1 and y_2 that correspond to n_1 and n_2, but others which correspond to $n_1 - \frac{1}{2}$ and $n_2 + \frac{1}{2}$ instead. From the general relation (101) which exists between y and n we easily find these limits, and write down the formula

$$\sum_{n=n_1}^{n_2} P_m(n) \doteq \int_{y_1 - \frac{1}{2\sqrt{mp(1-p)}}}^{y_2 + \frac{1}{2\sqrt{mp(1-p)}}} P_m(y)dy. \qquad (125)$$

We have chosen to speak of $P_m(y)$ as defined by the series (124), but the validity of our argument would be just exactly the same if we were to use some other definition. For instance, we might use the series (102) which is, except for the form in which it is expressed, identical with (124). But though such a substitution does not impair the validity of the argument, it does affect the ease with which computations can be carried out.

To illustrate this, we may note that if the series (102) were integrated it would give another series likewise containing various powers of y. To evaluate (125), therefore, we should have to compute these powers for both limits of integration and form the appropriate sums. On the other hand, the fact that the various ϕ's in (124) are successive derivatives of the Normal Law makes it only necessary to reduce the indices by 1 in order to have their integrals, the integral of $\phi'''(y)$ being $\phi''(y)$, the integral of $\phi^{iv}(y)$ being $\phi'''(y)$, and so on. Now all these quantities may be taken directly from the tables contained in Appendix V, even the integral of $\phi(y)$ being given under the heading $\phi_{-1}(y)$. It therefore becomes an exceedingly simple matter to carry out the computation of (125).

As a numerical example we might consider the problem of finding the chance of n successes in 100 trials of an event the probability of which is 0.1. For this problem we have $p = 0.1$, $m = 100$, $\sigma = 3$, whence (124) becomes

$$P_{100}(y) = \phi - \frac{2}{45}\phi''' + \left(\frac{23}{10{,}800}\phi^{iv} + \frac{2}{2{,}025}\phi^{vi}\right) + \left(\frac{1}{50{,}625}\phi^{v} - \frac{23}{243{,}000}\phi^{vii} - \frac{4}{273{,}375}\phi^{ix}\right) + \ldots . \qquad (126)$$

Table XVII illustrates the degree of accuracy which may be expected of this formula; but the form of the table requires some explanation.

TABLE XVII

THE BINOMIAL LAW $C_n^{100}\ (0.1)^n\ (0.9)^{100-n}$ AND SEVERAL GRAM-CHARLIER APPROXIMATIONS TO IT

n	True Value	Δ_0	Δ_1	Δ_2	Δ_3
0	0.00003	+0.00048	−0.00013	−0.00006	−0.00002
1	0.00030	+ 118	− 0	− 5	− 2
2	0.00162	+ 218	+ 32	+ 6	+ 1
3	0.00589	+ 285	+ 64	+ 18	+ 5
4	0.01587	+ 212	+ 52	+ 14	+ 3
5	0.03387	− 71	− 16	− 7	− 3
6	0.05958	− 491	− 95	− 22	− 5
7	0.08890	− 824	− 107	− 14	− 1
8	0.11482	− 834	− 31	+ 4	− 1
9	0.13042	− 462	+ 76	+ 12	+ 1
10	0.13187	+ 111	+ 111	− 1	− 1
11	0.11988	+ 592	+ 53	− 14	+ 1
12	0.09879	+ 770	− 33	− 2	+ 2
13	0.07430	+ 635	− 82	+ 12	− 1
14	0.05130	+ 370	− 59	+ 14	− 3
15	0.03268	+ 48	− 7	+ 2	− 1
16	0.01929	− 130	+ 30	− 8	+ 2
17	0.01059	− 185	+ 37	− 9	+ 3
18	0.00543	− 163	+ 22	− 4	+ 1
19	0.00260	− 113	+ 6	+ 1	− 1
20	0.00117	− 66	− 4	+ 3	− 1
21	0.00050	− 34	− 6	+ 2	− 0
22	0.00020	−0.00015	−0.00005	+0.00001	+0.00000

In the first place, (126) gives the values of $P_{100}(y)$, not $P_{100}(n)$. To get the latter it is necessary to introduce the proper Jacobian, which we know to be $1/\sigma = \frac{1}{3}$. Hence the values from which the table is derived are not the numbers given by (126), but one-third of them instead.

In the second place, though we have given in the second column the true values of the Binomial Law, the Gram–Charlier approximations to it are indicated, not by their actual values, but by their errors; for it is the errors in which we are principally interested.

In the third place, the four columns of deviations correspond:

(1) to the Normal Law, which is represented by the first term of of (124); (2) to the second approximation obtained by including the term which has the first power of σ in its denominator; (3) the approximation obtained by including the second power of σ also, and (4) the complete expression, so far as we have written it. That the convergence of the series is a fairly rapid one is obvious from the rapid vanishing of the Δ's.[1]

---

[1] To these remarks we may also add the following observations, which will probably be of less interest to the elementary student than to those already familiar with the subject.

In the first place, various writers, of whom Edgeworth is perhaps the best known, have pointed out that to get the "best" results from a Gram-Charlier Series the terms should be associated in an order different from the "natural" one. Specifically, the term of zero order must either be used alone, or else in some one of the combinations

$$0,\ 3;$$
$$0,\ 3,\ 4,\ 6;$$
$$0,\ 3,\ 4,\ 6,\ 5,\ 7,\ 9.$$

Now this is just exactly the way the terms have grouped themselves in (124) and (126); and we see at once that the rule may be said to be a natural consequence (in the case of Binomial expansions at least) of the attempt to arrange the terms in descending powers of σ.

In the second place, there is a common rule to the effect that the coefficient of the term of order 6 is approximately equal to half the square of the coefficient of order 3. In the case of (124) and (126) this is identically true, if by the "term of order 6" we mean that one which occurs in combination with ϕ^{iv}. This, however, is not the entire term of order 6: the first *unwritten* term of (124) would also contain ϕ^{vi}, but as it has a higher power of σ in its denominator it can be expected to be of little importance by comparison with the part accounted for by the common rule.

It would appear, therefore, that to obtain the best results from the use of the Gram-Charlier Series, *some* other order than the natural one is required. Whether that suggested by Edgeworth is the "best in the long run" will depend largely upon what we mean by that phrase. Certainly it is not to be expected that any order of summation can ever be devised which will not, in an exceptional case, be less exact than some other order which happened to be peculiarly appropriate to that exceptional case.

As an illustration of the fact that the grouping of the terms really has a beneficial effect at times, we may compare our approximations with that obtained by Mr. Arne Fisher in his *Mathematical Theory of Probabilities*, p. 268. The approximation there given was obtained from the use of terms as high as ϕ^{iv}, in the natural order, and therefore occupies an intermediate position between our Δ_1 and Δ_2 columns; for our grouping of terms is such that we either omit the function of the fourth order entirely, or else include the sixth also. We can therefore best make our comparison with these two columns.

Taking the root mean square of the deviations as our criterion, Mr. Fisher's approx-

We may also make use of the same illustration as an example of the type of accuracy that can be attained when the integrals of such series are used as approximate expressions for the sum of the Binomial Series itself. Table XVIII contains, in the second column, the results of summing the Binomial Series from n_1 to 100: that is, it contains the probability that 100 trials would result in no less than n_1 successes. The remaining columns contain the results of integrating (126) from the lower limit that corresponds to $n_1 - \frac{1}{2}$ $\left(\text{that is, from } y_1 - \frac{1}{2\sigma}\right)$ to infinity. The approximation is not quite so good as in Table XVII, but it would still be quite sufficient for many purposes.

---

imation is in error by 8 in the fourth decimal place (his values are listed to only four places), while our second approximation is in error by 5, and our third by 1, in the same place. In other words, a suitable grouping of the terms has given noticeably greater accuracy to our second approximation than was obtained from Mr. Fisher's use of the natural order *which included one term more.*

Mr. Fisher is, of course, well aware of the benefit to be derived from proper grouping. In fact, in a letter to the editor of the *Bell System Technical Journal* (Vol. 6, 1927, pp. 172–180), he has pointed out that there exists a definite relationship between the order of magnitude of the so-called "elementary errors" and the arrangement of the terms in the Gram-Charlier Series. Mr. Fisher informs me that the first 7 approximations are contained in the following combinations, originally given by Charlier and Jørgensen:

0;
0, 3;
0, 3, 4, 6;
0, 3, 4, 6, 5, 7, 9;
0, 3, 4, 6, 5, 7, 9, 8, 10, 12;
0, 3, 4, 6, 5, 7, 9, 8, 10, 12, 11, 13, 15;
0, 3, 4, 6, 5, 7, 9, 8, 10, 12, 11, 13, 15, 14, 16, 18.

These differ from the groupings to which we have been led only in the matter of including the entire term of any order in the place where that order first appears, whereas our scheme leads us to split such terms up into various parts according to the relative powers of σ which they contain. For example, in our seventh approximation the order of terms would be

0; 3; 4, 6; 5, 7, 9; 6, 8, 10, 12; 7, 9, 11, 13, 15; 8, 10, 12, 14, 16, 18;

the semicolons being used to mark off groups which have the same power of σ. The law of formation of either set is obvious.

In his letter to the *Bell System Technical Journal* Mr. Fisher makes use of the third of the above relationships and adds a term of order 6 to the values given in his *Mathematical Theory of Probabilities.* The result thus obtained is identical with the Δ_2 of Table XVII.

We can carry out a similar line of argument with respect to the series (116), which was obtained as an approximation to the Poisson Law. It is found without difficulty to reduce to the Gram–Charlier form

$$P(y) = \phi - \frac{1}{\sigma}\frac{\phi'''}{6} + \frac{1}{\sigma^2}\left(\frac{\phi^{iv}}{24} + \frac{\phi^{vi}}{72}\right) - \frac{1}{\sigma^3}\left(\frac{\phi^{v}}{120} + \frac{\phi^{viii}}{144} + \frac{\phi^{ix}}{1{,}296}\right) + \frac{1}{\sigma^4}\left(\frac{\phi^{vi}}{720} + \frac{13\phi^{viii}}{5{,}760} + \frac{\phi^{x}}{1{,}728} + \frac{\phi^{xii}}{31{,}104}\right) + \ldots,$$

from which either the probability of n taking a particular value or the chance of it lying within a specified range could be computed. The Poisson Law is of sufficient importance, however, that extensive

TABLE XVIII

CERTAIN GRAM-CHARLIER APPROXIMATIONS TO THE SUM OF A BINOMIAL SERIES

n_1	True Value	Δ_0	Δ_1	Δ_2	Δ_3
0	1.00000	−0.00023	+0.00020	+0.00002	−0.00001
5	0.97629	− 967	− 187	− 99	− 81
10	0.54871	+ 1747	+ 47	− 7	− 6
15	0.07257	− 576	+ 143	+ 65	+ 80
20	0.00198	−0.00121	−0.00015	+0.00011	+0.00009

tables have been prepared, not only of $P(n) = \epsilon^n e^{-\epsilon}/n!$ itself, but also of the function $\Pi(n) = \sum_n^\infty P(n)$ which represents the probability that n is not less than a specified value. Appendices VI and VII are skeleton tables of this sort. Much more elaborate ones can be obtained, two of which are referred to in the References for Outside Reading. Obviously, with such tables available, we have very little use for a series expansion.

§ 94. *Empirical Families of Curves; Transformation of Variable*

Let us think for the moment of any two distribution curves, such as those shown in Fig. 33. They are decidedly dissimilar, but being distribution functions they necessarily enclose equal areas. Suppose, now, that we start from the left and lay off on each an equal area A, as, for example, the shaded areas in the figures. Finally, suppose we call the y and x which

bound these areas "corresponding" values. Such a process could obviously be carried out for any area A (or for any x), thus relating a y to every x. Let us imagine the corresponding

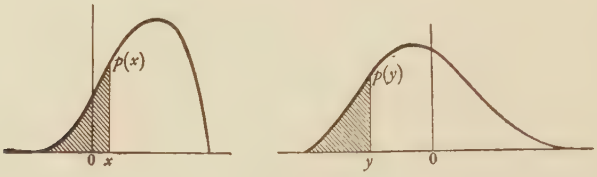


FIG. 33.

pairs to be plotted as in Fig. 34. By this means we have found a function of x the distribution of which is exactly in accordance with the right-hand curve of Fig. 33.

In other words, *no matter what the distribution of x may be, there is some function y(x) of such a nature that the distribution of y will conform to any law we may desire.* For example, we might cause it to be represented by a straight line, so that equal ranges of y were equally probable.

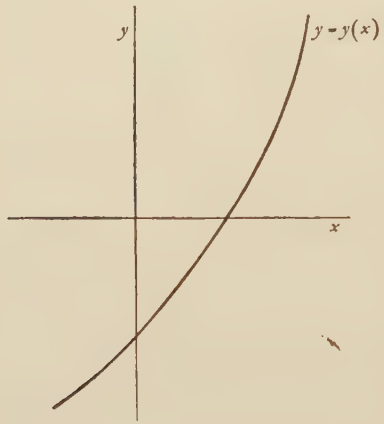


FIG. 34.

Little has ever been done in the matter of finding to what extent this form of transformation might be utilized in deriving distribution laws for empirical data,[1] but it has been used now and then to throw unusual types of distribution curves into forms that conformed better to the established empirical types.

In a way this too is a means of obtaining a family of curves,

[1] It would seem that there might be some point to the attempt to treat statistical data by finding what transformation would result in a completely random distribution: that is, in equal probabilities for all intervals within a certain range.

but in this instance the family includes almost anything conceivable. We have, in fact, progressed from a few types, derived from a given differential equation and containing only a limited number of parameters, through a somewhat more general set with more parameters and therefore theoretically capable of finer gradations, to a method which, in its broad outlines at least, is capable of yielding anything at all which can conceivably be a distribution curve. But this progression to greater and greater flexibility has been accompanied by a loss of definiteness of type which in practice seems to largely nullify any value it might otherwise possess. After all, statistical data is not *exact*. It is merely *representative*, and about all that we can hope to do with it — if we may be allowed for the sake of emphasis to use an expression which smacks somewhat of exaggeration — is to find the type to which it belongs.

PROBLEMS

1. Show that Poisson's Law leads to the differential equation (117).

2. Find the value of k in (118).

3. Find the first two expectations of $(x - \alpha)$ in (118).

4. Find a general formula for the ith expectation of $(x - \alpha)$ in (121).

5. In the case of (121) let δ be the deviation of x from its expectation, measured in terms of its standard deviation. Write the equation of $P(\delta)$.

References for Outside Reading

I. *On the Normal Law:*

1. Coolidge: *Probability*, pp. 101–119.
2. Lévy: *Calcul des Probabilités*, Part I, Chapter IV, and Part II, Chapters IV and V.
3. Czuber: *Wahrscheinlichkeitsrechnung*, pp. 287–299.

II. *Pearson's Curves:*

4. Pearson: *Mathematical Contributions to the Theory of Evolution*, Philosophical Transactions, A, Vol. 186 (1895), pp. 343–414; Vol. 197 (1901), pp. 443–459; Vol. 216 (1916), pp. 429–457.

III. *Gram–Charlier Series:*

5. Arne Fisher: *The Mathematical Theory of Probabilities*, Chapters XIII and XIV.

IV. *General:*

6. Rietz: *Report on Statistics*, Bulletin of American Mathematical Society, Vol. 30 (1924), pp. 416–453; *Mathematical Statistics* (No. 3 in the series of Carus Mathematical Monographs).

V. *Tables:*

7. Pearson: *Tables for Biometricians and Statisticians*, Cambridge University Press.
8. Jørgensen: *Mathematical-Statistical Tables*, D. Van Nostrand, New York.
9. Glover: *Tables of Applied Mathematics in Finance, Insurance and Statistics*, Wahr Publishing Co., Ann Arbor, Mich.

CHAPTER IX

Curve Fitting

§ 95. *The Primary Problem*

We have said that many probabilities cannot be found by the application of the definition of a probability, and that in such cases experiment is our best guide. We have also seen, in our discussion of Bernoulli's Theorem, that the most we can conclude from experiment is that the results are *unlikely* to differ *much* from the true probability. Naturally, we are interested to know *how unlikely* and *how much*. The two things are of course very intimately related — there is a certain probability of deviating from the truth by any mentionable amount — and thus the problem that presents itself is to find the relation between these two.

When we can get the necessary information, Bayes' Theorem tells us how sure we may be of the estimate to which an experiment has led us. But it requires that we know how likely that same estimate was before the experiment was performed, and we do not usually possess that information. Hence Bayes' Theorem is seldom capable of assisting us. What we really need is some method of judging the accuracy of our estimate without knowing what its *inherent* probability was before the experiment was performed. The present chapter, in the main, is devoted to methods which have been developed for this purpose.

From the very start, however, we must keep in mind this thing: It is a *merit* of Bayes' Theorem, not a *weakness*, that it takes account of the inherent likelihood of an event. It would be well to read again § 52 and note that the information sought does not depend solely on the experimental evidence,

but on all the other bits of knowledge which go into the making up of our advance judgment as to the goodness of our coin.

It is a matter of cold fact that the reliance which we dare place in the result of an experiment depends upon the inherent plausibility of the result as well as upon the accuracy with which it has been attained. Any substitute for Bayes' Theorem which does not require the knowledge of the a priori *probabilities must therefore be incapable of giving the* a posteriori *probabilities.*

Being, therefore, unable to get a mathematical measure of the assurance with which we may accept our estimate because we do not first possess a mathematical measure of its inherent plausibility, we turn to the task of finding the best possible makeshift. Fortunately the makeshift is not a very unsatisfactory one when its meaning is clearly understood, but we must repeat *that it does not tell us how probable our estimate is. Nothing can tell us that which does not require the* a priori *probabilities.*

In entering upon this subject we are embarking upon the study of statistical methods. We cannot cover that study completely, for the mere description and illustration of the processes of computation alone would occupy an entire book, and an adequate treatment of the algebraic discussions upon which those methods are founded would occupy several. Our purpose must be, first, to give a correct picture of what we may expect to learn from such statistical studies; second, to direct the student to the sources where the detail can best be acquired when needed; and third, to explain the more important technical terms employed, so that he will not find those sources unintelligible.

§ 96. *The Hypothetical "Universe of Data," or "Population," and the "Sample"*

When an "experiment" is performed we really sort out one of a group of possibilities. Thus, when a penny turns up "heads" we sort out one of *two* possibilities; or when we

measure the resistance of an electrical device we sort out one of an infinity of possibilities. I do not mean to call attention to the fact that the device might really have most any resistance, and that we are determining what it really is. I mean that, due to the uncertainties of measurement, we *might* have gotten some other answer: in other words, that our mistake in measurement is only one of a number of possible mistakes. Theoretically we may take the number of possibilities to be infinite, though ordinarily, due to the fact that our instruments have a "least count," this will not be true.

When a punch press turns out a stamping, its size is one of a number of sizes that *might* have appeared.

When a man dies, his age is one of those to which he *might* have lived.

Each of these is, in a way, an "experiment." But experiments need not be so simple. The day's output of a punch press is — so far as dimensions are concerned — an aggregate of data: a table of numbers. This table, however, is only one of a number of tables that *might* have been obtained. It, too, is an "experiment" that could have resulted otherwise.

The number of heads and tails that come from n throws of a penny is only one of a number of possible results. And so on.

Thinking along these lines leads us quite naturally to the conception of a vast storehouse of possibilities from which, by our experiment, we have drawn one blindly. This hypothetical storehouse we may call the "universe" (or "population") from which our "sample" was drawn. Naturally, we must imagine it to contain all possible results in proportions equal to their respective probabilities, which will ordinarily require us to think of an infinity of items in the storehouse (or, "individuals" in the "population"). Moreover, as we have said, the individual item may be quite complicated: it may be a sort of bundle of things. Thus each individual in the population may be "n heads and $12 - n$ tails," if we are thinking of what can happen in twelve throws of a penny.

These words "universe" or "population," "individual,"

and " sample " are in constant use in statistical literature, and they are really quite valuable, for they enable us to talk quite abstractly about events of such diverse sorts that no sufficiently general word could otherwise be found for them. We have introduced them as relating to a hypothetical storehouse the properties of which are determined by the probabilities, known or unknown, which govern our experiment. It is only fair to say that they did not so originate. They are the product of a statistical school which, if it does not actually deny that a probability exists before an experiment is performed, at least regards the experiment as the only feasible means of arriving at it. If I interpret them correctly, the adherents of this school attribute to a " population " a degree of logical reality which transcends that of abstract " probability." To them, the " population " or " universe " is the actuality, while " probability " is relegated to a shorthand form of statement for " the relative frequency of like individuals in the population."

Though I appreciate the practical usefulness of the " universe " concept in aiding imagination, I do not see that the logic of a subject is in any way bettered by substituting for an abstraction (probability) that sort of quasi-concreteness which dreams possess: that is, the mental image of a " universe " which would be concrete enough if it existed, but which never does, and frequently could not, exist.

§ 97. *The Accepted Criterion as to Goodness of Fit*

We may now state the form of solution which is adopted as a standard substitute for Bayes' Theorem. It consists of

(*a*) Making use of our experimental data to form a more or less plausible estimate of the probabilities which the experiment was designed to measure; and

(*b*) Computing the chance that another experiment, so conducted that its probabilities were really equal to these estimates, would lead to a result that is at least as improbable as the one under discussion.

Or we may express the matter in terms which are more commonly used by statisticians as follows:

(*a*) We consider the results of our experiment and from them estimate the proportions in which the different kinds of individuals are contained in the population; and

(*b*) We compute the chance that another sample, taken from this assumed population, would contain the various kinds of individuals in proportions that are no more probable than those in the experimental sample under discussion.

A simple illustration will help in getting these ideas fixed. Suppose we were attempting to measure the probability of a certain event occurring, and had conducted an experiment which gave 20 successes in 50 trials. Suppose, moreover, that because of certain collateral evidence, which we need not discuss, we reached the conclusion that a fair estimate of the probability was $\frac{1}{3}$. From Table X we find that *if this estimate is correct* the chance of getting 20 successes in 50 trials is only 0.0704. The particular result in question is therefore not very likely. But by glancing over the table we quickly observe that no other result is much more likely, and many of them are even less likely. Obviously under these circumstances it is unfair to conclude that our estimate is an unreasonable one. The question that really presents itself is, how likely would we be to get a result the probability of which is no greater than 0.0704. We can easily answer this question by adding together the probabilities of all possible results except those from 14 to 19, thus getting the answer 0.368. Obviously, then, the chance of getting a result that was at least as unlikely as 20 successes is big enough that we need not discard our assumed probability as an untenable one.

Or, to take another simple example, suppose we have dealt a pack of cards and have observed the order in which they fell. Suppose by *a priori* reasoning we have formed the estimate that all orders are equally likely. On the basis of this estimate we compute the chance of getting the observed result and find it to be $1/52!$, which is inconceivably small. Are we therefore justified in discarding our hypothesis? Certainly not, for any

other possible order would be just as improbable. The thing that counts is, that on the basis of our hypothesis we were *sure* to get as unlikely a result as this.

The accepted form of solution tells us, then, not how likely it is that our estimates are correct, nor how likely our experimental result was if the estimates are correct, but instead it tells us the chance of getting a result at least as unusual as the one we obtained, if the estimates are correct.

The first step in our process — that of forming an estimate of our probabilities — is likely to present itself in a wide variety of forms. Sometimes it requires the estimation of only one probability, as in the first example above; sometimes it requires the estimation of many, as in the card illustration where we really estimated 52! probabilities at once, making them all equal; sometimes the number may be infinite, as in finding the chance of a metal stamping differing from standard by a specified amount. The last case, indeed, requires that we give the equation of a distribution curve, rather than the values of certain discrete numbers.

Naturally the methods of forming our estimates of these probabilities are going to vary widely, not only with the number of variables with which we must deal, but also with the amount of collateral information which we may possess and which influences our choice. For this part of our work the various distribution functions discussed in Chapter VIII will be our principal stock of tools.

Once our estimate has been formed, however, the other half of the process follows the same set of ideas pretty consistently, no matter whether we may be dealing with a single probability or with an entire distribution curve. It will be our purpose, so far as possible, to cause this underlying unity to show through the confusion of algebraic detail which we cannot entirely avoid. That purpose can probably best be accomplished by means of a set of examples arranged in increasing order of complexity.

§ 98. *Some Instructive Illustrations; The Biased Die*

EXAMPLE 49.—*There is a case on record where a die was thrown 315,672 times, with the result that either 5 or 6 appeared 106,602 times. Was the die true, and, if not, what was the probability of the appearance of 5 or 6?*

If the die was a true die the expectation was just one-third the number of throws, or 105,224. The difference is not very great — about $1\frac{1}{3}$ per cent — certainly not great enough to convince us offhand that the die was bad. We therefore begin with the assumption that it was *good*, and attempt to find out whether that assumption is plausible. To this end we shall apply our criterion of goodness of fit, and shall compute the probability than an experiment *conducted with a true die* would give a result that is at least as unlikely as 106,602 5's or 6's.

This is not very difficult. In fact, the case is one to which the Binomial Law applies directly, the number of trials being 315,672 and the (assumed) probability $\frac{1}{3}$. But with so large a number of trials the Binomial Law is known to be very accurately approximated by the Normal Law, provided we choose our unit of measure properly. That is, we must first find the deviation of our result from expectation, which is obviously 1,378, and then we must express it in terms of the standard deviation as a unit. We have seen in § 82, however, that the standard deviation of the Binomial Law is $\sqrt{mp(1-p)}$, which in this case is found to be 264.9. Measured in terms of this unit the deviation from expectation is $y = 5.20$. Obviously, larger deviations than this are less probable and smaller ones more probable; so our problem becomes that of finding the chance of a deviation at least as big as 5.20.

This is identically the same problem as was met in § 93, except that in the present instance no very high degree of accuracy is needed, so that the first term of the series (124), which is just the Normal Law, is quite sufficient. Moreover we can quite properly ignore the small correction term

$1\ 2\sqrt{mp(1-p)}$ that appears in the limits of the integral (125). Hence, with accuracy enough for our present purposes, we may say that the chance of a deviation larger than 5.20 is

$$\frac{1}{\sqrt{2\pi}}\int_{5.20}^{\infty} e^{\frac{-y^2}{2}}\,dy = \phi_{-1}(\infty) - \phi_{-1}(5.20),$$

where $\phi_{-1}(y)$, as in § 93, refers to the function tabulated in Appendix V.

We have thus taken care of deviations which are positive and greater than 5.20. But to the extent to which the Normal Law is an approximation to the Binomial there are also negative deviations that have just the same probabilities. Hence the total chance of getting a result less likely than that stated in the example is just twice what we have written above.

We could, of course, compute this result by the use of Appendix V. However, the integral

$$\Phi(y) = \frac{2}{\sqrt{2\pi}}\int_{y}^{\infty} e^{\frac{-y^2}{2}}\,dy \qquad (127)$$

is of so frequent occurrence that it is worthy of a table of its own, which is given in Appendix IV; and from this table we find at once that the probability in question is $\Phi(5.20)$, which is obviously less than 0.0001. From a larger table its value could be found to be 0.000,000,2. Hence *the chance of a true die giving as improbable a result as did our experiment is only about one in five million.* It is therefore quite likely that the die is asymmetrical *unless there is a very powerful* a priori *reason for thinking otherwise.* As we have no such powerful prejudice, we accept our first estimate of p to be unlikely, and seek for a better one.

Suppose we adopt, as our second estimate of the probability in question, the proportion of times a 5 or 6 was observed, for we know from Bernoulli's Theorem that this is not likely to differ much from the true value of p. This gives us $p = 0.337,699$. But upon this assumption we find that $y = 0$. Naturally any experiment which we might perform would

show a deviation as large as this; or in other words $\Phi(y)$ is, for this case, unity. But surely this does not mean that our new estimate is absolutely accurate. Instead it merely reflects the fact that we have artificially forced an agreement between the estimate and the experiment by computing the one from the other.

Later on, in dealing with more complicated examples, we shall find that we must always compensate for any such forcing and it will be well for the student to keep in mind until then the present illustration, in which the absurdity of accepting the uncompensated $\Phi(y)$ is quite apparent. The very simplicity

TABLE XIX

Assumed p	Expected Number of 5's or 6's	Deviation	Standard Deviation	y	$\Phi(y)$
0.333	105,119	−1483	264.9	−5.699	0.0000
0.335	105,750	+ 852	265.2	+3.213	0.0013
0.337	106,382	+ 220	265.6	+0.828	0.4076
0.339	107,013	− 411	266.0	−1.545	0.1224
0.341	107,645	−1043	266.4	−3.915	0.0001

of our example, however, makes the present an unsuitable time to discuss the matter. Instead we shall circumvent the difficulty by another method of approach.

We wish to know how far we can trust this experimental approximation to the probability. Surely we can place no confidence in the last 9 of the 0.337,699, for if we had tossed the die just once more we would of necessity have had a frequency ratio of either 0.337,696 or 0.337,702. On the other hand, to write merely 0.33 would be unduly conservative, for we have seen that the departure from 0.333,333 is almost certainly significant, and that departure occurs only in the third place.

We answer this question by assuming various probabilities to be the true one — just as we have already assumed first $\frac{1}{3}$ and then 0.337,699 to be the true one — and finding whether

the deviations of the experimental results from these assumed probabilities can be regarded as significant. We list the computations involved in Table XIX.

It is obvious that the results which we have obtained were to be expected if the probability is 0.337 or 0.339, but that they would be quite unusual for $p = 0.335$, and exceedingly so for the other values. Hence, unless there is a powerful reason for doubting it, we are forced to conclude that p is very probably between 0.335 and 0.340.

This is one way of answering our question, and, indeed, virtually the only one open to us; but we have made the computations unduly laborious, as we shall readily see by making two observations about the above figures. In the first place, in looking up the probabilities in Appendix IV we have made no use of the sign of y. Therefore y's of equal magnitude but opposite sign are equally likely — or we should rather say, the deviations to which they correspond are equally likely. Why, then, can we not choose that Φ which we shall accept as marking the boundary between admissible and inadmissible p's, look up the corresponding y, and find from it, first the deviations, and then the p's, which we are to regard as limiting our range? That is, why can we not work *backwards* from Φ to the corresponding p?

The hypercritical answer is, that to get the deviations from the y we should have to multiply y by certain standard deviations and specifically by those standard deviations which correspond to the p's for which we seek, and which are therefore as yet unknown. But a second observation here comes in to save us from this difficulty: the standard deviations are almost identical for all the p's which we have considered. It therefore makes very little difference whether we use one of them or the other. In particular, we could quite properly use one computed from the observed ratio, 0.337,699.

Let us carry out this computation.

Let us agree to admit as plausible any value of p for which $\Phi(y)$ exceeds 0.01. From Appendix IV we find the corresponding value of y to be 2.576. Next we compute an approximate

standard deviation by using for p the experimental approximation 0.3377. The result is 265.7. Multiplying this by $y = 2.576$ we find that our allowable deviations are ± 684. This means that the limiting expectations are $106{,}602 \pm 684$, and that therefore the limiting p's are $(106{,}602 \pm 684)/315{,}672 = 0.3377 \pm 0.0022$.

If, then, p is less than 0.3355 or greater than 0.3399 there is less chance than one in a hundred that an attempt to repeat the experiment would give a result that differed from expectation by as much as did the one under discussion.

We could even reduce the labor still farther by choosing in place of 0.01 as the particular probability which we will begin to regard as "unusual," that probability which corresponds to $y = 1$. (It is actually 0.3176.) We thus save ourselves the necessity of finding y from Appendix IV, and we also save a multiplication, getting our $\pm$ term by simply dividing the standard deviation by the number of trials. In our example this gives $265.7/315{,}672 = 0.000{,}84$.

This last is the usual practice, the answer to the problem being written $0.3377 \pm 0.000{,}84$. Without any further explanation the reader then understands that this $\pm$ term corresponds to $y = 1$, and forms his judgments accordingly. In particular, if he wishes to set such limits upon p that, if p is really outside them, there is less chance than 0.001 of getting a sample which deviates as much from expectation as did his experiment, he would look up in Appendix IV the value 3.291 which corresponds to $\Phi = 0.001$ and multiply the $\pm$ term by it, thus getting as his limits $p = 0.3377 \pm 0.0027$.

§ 99. *Discussion of Example 49*

There are several things that should be said about the arguments which have been used in the last section.

To begin with, it should be noted that we have done two things that are, in a sense at least, quite distinct. First, we attempted to answer the questions, "Is $\frac{1}{3}$ a sensible value for p?" and "Is 0.337,699 a sensible value for p?" Then we

attempted to answer the question, "Within what limits have we succeeded, by our experiment, in confining p?" The first question asks about the legitimacy of our guesses, the second asks about the precision of our experiment; and the elaboration of each of these points of view leads to a branch of the Theory of Probability which is of quite sufficient breadth to warrant its treatment as a separate subject of study. The first leads to generalized concepts regarding Goodness of Fit; the other to the subject of Precision of Statistics.

We shall not in this text be concerned with the subject of Precision of Measurement; but having introduced this simple example, it is wise for us to note particularly the basis on which we justified its several steps. First, we felt assured that we were dealing with a case of independent trials under identical conditions; and that therefore our data obeyed the Binomial Law. The Normal Law entered only as an approximation to the Binomial. In § 82, in discussing such uses of the Normal Law, we saw that the approximation was fair so long as y^3/σ was small. This would seem to justify its use in the present case since y^3/σ is less than 0.7 in all of our computations. Second, we made use of the fact that the standard deviation was approximately constant over the range of values of p which interested us, justifying this by the fact that we had already computed enough standard deviations in Table XIX to know that it was true.

It is entirely possible, however, to set up illustrations in which these approximations are not so well justified, and hence it is unwise to make use of such methods without keeping in mind the conditions which underlie them. The student to whom the subject of Precision of Statistics is of interest should follow the matter further in some treatise on the subject, such as those given at the end of the chapter under the References for Outside Reading.

Finally, with respect to the first half of our discussion, in which we sought to determine whether such a guess as $p = \frac{1}{3}$ or $p = 0.337{,}699$ was tenable, we must again repeat that a high value of Φ is only significant (1) provided this high value

of Φ has not been artificially brought about through the use of our data in determining p, and (2) provided there is no *a priori* prejudice against the particular value of p which is being considered. As we proceed with the chapter we shall learn how to compensate for the bias that is brought about by the use of our data in the computation of our probabilities; but we can do nothing whatever to take account of their inherent likelihood. If, therefore, this inherent likelihood is to receive any consideration at all, it must be in the mind of the investigator when he comes to the interpretation of his results.

§ 100. *Some Instructive Illustrations; Weldon's Dice Data*

The figures which we gave in Example 49 were totals obtained from an experiment made by the English biologist, Weldon, who performed it under circumstances somewhat different from those stated in the example. Instead of using one die, he used twelve, throwing all twelve at once and recording the number of 5's and 6's that appeared. Table XX contains a summary of his results. We shall now turn our attention to the data in this form, considering in particular the two questions which follow:

EXAMPLE 50.—*Is Weldon's dice data, as contained in Table XX, consistent with the assumption that the twelve dice were unbiased?*

EXAMPLE 51.—*Is Weldon's dice data, as contained in Table XX, consistent with the assumption that the probability of 5 or 6 had the same value 0.3377 for each of the twelve dice?* [1]

---

[1] If we admit that the dice are very probably all different, and denote the probabilities of showing 5 or 6 by $p', p'', \ldots, p^{xii}$, we can so determine these p's that all of Weldon's observed values agree with their expectations. This is exactly analogous to what we did in the latter half of § 98. Before we could place much reliance in our results, however, we should be forced to find within what limits each of these values would have to be confined if the actual observations were not to be too improbable. That is, we should have to determine the precision of this experimental determination of the various p's.

This procedure is probably a more logical one to follow than either of those outlined

We shall find that it is best to carry these two examples through simultaneously, as most of our remarks about one of them will be equally true of the other also.

TABLE XX

WELDON'S DICE DATA

Number of 5's or 6's	Observed Frequency	Frequency Ratio
0	185	0.007033
1	1,149	0.043678
2	3,265	0.124116
3	5,475	0.208127
4	6,114	0.232418
5	5,194	0.197445
6	3,067	0.116589
7	1,331	0.050597
8	403	0.015320
9	105	0.003991
10	14	0.000532
11	4	0.000152
12	0	0.000000
	Total 26,306	

We note, to begin with, that when the twelve dice are thrown the chance that just j show 5's or 6's is, in either case, given by the Binomial Law [1]

$$p_j = C_j^{12}\, p^j\, (1 - p)^{12-j};$$

the only difference being that the p is $\frac{1}{3}$ in one case and 0.3377

in the examples, but unfortunately it requires a great deal of computation in finding the first estimates of the p's, and requires a more complicated form of statement regarding the limits to which they must be confined. In fact, it is not a practical method at all, unless the problem justifies considerable labor.

The questions which we have formulated are, on the other hand, not at all difficult to deal with from a computational standpoint, though they lead to some theoretical considerations which are not altogether simple.

[1] We would ordinarily write $P(j)$ instead of p_j, but the uses which we are to make of the symbol in the present case are such that it is desirable to use the simpler symbol.

in the other. The second columns of Tables XXI and XXII are obtained by direct computation from this formula.

TABLE XXI

Discussion of Weldon's Dice Data if Dice were True

Number of 5's and 6's	Probability	Observed Frequency	Expected Frequency	Deviation from Expectation	Divergence
j	p_j	n_j	$\epsilon(n_j)$	δ_j	$\frac{\delta_j^2}{\epsilon(n_j)}$
0	0.007707	185	202.75	− 17.75	1.554
1	0.046244	1,149	1,216.50	− 67.50	3.745
2	0.127171	3,265	3,345.37	− 80.37	1.931
3	0.211952	5,475	5,575.61	−100.61	1.815
4	0.238446	6,114	6,272.56	−158.56	4.008
5	0.190757	5,194	5,018.05	+175.95	6.169
6	0.111275	3,067	2,927.20	+139.80	6.677
7	0.047689	1,331	1,254.51	+ 76.49	4.664
8	0.014903	403	392.04	+ 10.96	0.306
9	0.003312	105	87.12	+ 17.88	3.670
10	0.000497	14	13.07	+ 0.93	0.066
11	0.000045	4	1.19	+ 2.81	6.143
12	0.000002	0	0.05	− 0.05	
			26,306.02		$\chi^2 = 40.784$

Next we must remember that each of Weldon's 26,306 throws constituted an independent trial, and that therefore the expected number of throws showing just j 5's or 6's is $26{,}306\, p_j$. These expected frequencies are shown in the fourth columns of the tables. The third columns contain the observed frequencies, as taken directly from Table XX, while in the fifth columns are listed the deviations from expectation, δ_j.

This completes that part of our discussion of Examples 50 and 51 which concerns itself with estimating the magnitudes of the various probabilities. We must next attempt to find out how plausible the estimates are.

TABLE XXII

DISCUSSION OF WELDON'S DICE DATA IF DICE WERE EQUALLY BIASED

Number of 5's and 6's	Probability	Observed Frequency	Expected Frequency	Deviation from Expectation	Divergence
j	p_j	n_j	$\epsilon(n_j)$	δ_j	$\frac{\delta_j^2}{\epsilon(n_j)}$
0	0.007123	185	187.38	− 2.38	0.030
1	0.043584	1,149	1,146.51	+ 2.49	0.005
2	0.122225	3,265	3,215.24	+ 49.76	0.770
3	0.207736	5,475	5,464.70	+ 10.30	0.019
4	0.238324	6,114	6,269.35	−155.35	3.849
5	0.194429	5,194	5,144.65	+ 79.35	1.231
6	0.115660	3,067	3,042.54	+ 24.46	0.197
7	0.050549	1,331	1,329.73	+ 1.27	0.001
8	0.016109	403	423.76	− 20.76	1.017
9	0.003650	105	96.03	+ 8.97	0.838
10	0.000558	14	14.69	− 0.69	0.032
11	0.000052	4	1.36	+ 2.64	4.688
12	0.000002	0	0.06	− 0.06	
			26,306.00		$\chi^2 = 12.677$

§ 101. *An Approximation to the General Multinomial Law*

Viewed in a general way, Example 50 deals with a complete set of mutually exclusive events, for on a single throw of the twelve dice there must be either no successes, or one, or two, or some other number not exceeding twelve. Moreover, Weldon's 26,306 throws constitute repeated independent trials of exactly the sort contemplated in § 26. Hence the chance of each member of the set occurring with exactly a specified frequency can be determined from (24). In particular we could compute, if we desired, the chance that another 26,306 throws conducted under the conditions laid down in Example 50 would exactly reproduce Weldon's results. That, however, is not quite what we need. Instead we want to know the chance that such a test would give a result that is no more likely than

Weldon's, and this would seem to require the addition of a great many terms of the form (24). We must therefore seek for some simple approximation which will bear to the Multinomial Law (24) a relation analogous to that of the Normal to the Binomial.

Example 51 also deals with a complete set of mutually exclusive events, and to this extent is identical with 50. But the two differ in the same important respect that made the second discussion of Example 49 differ from the first: in 51 a certain degree of agreement has been forced upon Weldon's data by the act of computing the number 0.3377 from it. To be exact, we have forced an agreement between the average number of successes and the number expected, so that the relation

$$\frac{1}{m}\Sigma j n_j = \Sigma j p_j,$$

or, if we prefer to write it so,

$$\Sigma j n_j = m \Sigma j p_j, \tag{128}$$

is satisfied. It is therefore not fair to use, as a criterion for the plausibility of our assumed conditions, the unconditional probability of getting a result that is as unusual as Weldon's. Instead we must get the *conditional* probability that a test, conducted with the probabilities listed in Table XXII, *but required to satisfy the auxiliary condition* (*128*), would give a more unusual result than Weldon's.

Had we used a more elaborate formula than the Binomial in arriving at our estimates — for example, had we admitted the possibility of the dice not being identical — we might have forced an agreement between experiment and estimate in even more respects, and in that case our criterion for goodness of fit would be the *conditional* probability *if the hypothetical test were required to satisfy the same set of auxiliary conditions as were used in building up the estimate.* Hence for the purposes of our study we need, not only an approximation to (24), but also an approximation to a certain sort of conditional prob-

ability as well. We shall find, however, that the derivation of the conditional probability will be a very simple matter once the unconditional one has been obtained.

Let us, then, consider a set of s independent events, for which the individual probabilities of occurrence are $p_1, p_2, \ldots, p_s$. If m independent trials are made, the chance that the first of these events occurs just n_1 times, the second just n_2 times, and so on, is, by (24) and (5),

$$P_m(n_1, n_2, \ldots, n_s) = \frac{m!}{n_1!\, n_2! \ldots n_s!} p_1^{n_1} p_2^{n_2} \ldots p_s^{n_s}. \quad (129)$$

It must be remembered, of course, that

$$n_1 + n_2 + \ldots + n_s = m. \quad (130)$$

We now assume that every n_j is large enough that it can, without serious error, be replaced by its Stirling approximation.[1] Introducing these approximations, and making some simple rearrangements, (129) becomes

$$P = \frac{1}{(\sqrt{2\pi m})^{s-1}\sqrt{p_1 p_2 \ldots p_s}} \left(\frac{mp_1}{n_1}\right)^{n_1+\frac{1}{2}} \left(\frac{mp_2}{n_2}\right)^{n_2+\frac{1}{2}} \ldots \left(\frac{mp_s}{n_s}\right)^{n_s+\frac{1}{2}}. \quad (131)$$

Next we replace each n_j by the corresponding deviation measured in terms of its standard deviation. That is, we write

$$y_j = \frac{\delta_j}{\sigma_j} = \frac{n_j - mp_j}{\sigma_j},$$

where $\sigma_j^2 = mp_j(1 - p_j)$. This gives us, for each factor of (131), a new factor of the form

$$\left(1 + \frac{\sigma_j y_j}{mp_j}\right)^{-mp_j - \sigma_j y_j - \frac{1}{2}},$$

and we seek an approximate value of this when m is large.

We have carried out similar transformations often enough that there should be no need for extensive explanation. We

[1] This means that we shall not often want to use the resulting approximation when any n_j is less than 2.

take the logarithm of the entire group of factors, expand each term in a series, and collect the results according to descending powers of m. The final result is

$$\log\left[(\sqrt{2\pi m})^{s-1}\sqrt{p_1\, p_2 \ldots p_s}\, P\right]$$

$$= -\Sigma\left(\sigma_j\, y_j + \frac{\sigma_j^2\, y_j^2}{2\, mp_j} + \text{terms of lower order}\right). \quad (132)$$

Next we go back to (130) and replace all the n's by their corresponding y's. The result is

$$\Sigma\, \sigma_j\, y_j + m\, \Sigma\, p_j = m;$$

whence, since $\Sigma\, p_j = 1$, it follows that $\Sigma\, \sigma_j\, y_j = 0$. Therefore, when we remember the value of σ_j^2, (132) becomes

$$P(n_1, n_2, \ldots, n_s) = \frac{1}{(\sqrt{2\pi m})^{s-1}\sqrt{p_1\, p_2 \ldots p_s}}\, e^{-\frac{1}{2}\Sigma\, y_j^2(1-p_j)}. \quad (133)$$

The form of the exponent to which we are thus led suggests the substitution of a new variable $x_j = y_j\sqrt{1 - p_j}$ in place of y_j. A little later we shall discuss what this new variable signifies. For the moment we had best keep our attention focused upon the purely mathematical ideas. Let us, then, change to x_j in place of y_j, thus obtaining

$$P(n_1, n_2, \ldots, n_s) = \frac{1}{(\sqrt{2\pi m})^{s-1}\sqrt{p_1\, p_2 \ldots p_s}}\, e^{-\frac{1}{2}\Sigma\, x_j^2} \quad (134)$$

where

$$x_1 = \frac{n_1 - mp_1}{\sqrt{mp_1}},\ x_2 = \frac{n_2 - mp_2}{\sqrt{mp_2}},\ \ldots,\ x_s = \frac{n_s - mp_s}{\sqrt{mp_s}}. \quad (135)$$

This is the approximation to (129) for which we have been seeking. Before attempting to use it, let us take up the question, to which we referred at the beginning of the section, of the changes that must be made in it when our probabilities are subject to auxiliary conditions such as (128).

However many of these auxiliary conditions there may be, let us call the satisfying of them "event B." Then the thing which we desire to find is the probability of obtaining exactly

the values $n_1, n_2, \ldots, n_s$ *if event B occurs.* By Bayes' Theorem we have at once, however, the relation

$$P_B(n_1, n_2, \ldots, n_s) = \frac{P(n_1, n_2, \ldots, n_s)\, P_{n_1, n_2, \ldots, n_s}(B)}{\Sigma\, P(n_1, n_2, \ldots, n_s)\, P_{n_1, n_2, \ldots, n_s}(B)},$$

it being understood, of course, that the summation is to be taken over every possible set of values of the n's. If, then, we can find the conditional probabilities which occur on the right-hand side of the equation, we shall be able to get the formula for which we seek. This, however, is an exceedingly simple matter. A set of values of the n's either does or does not satisfy the auxiliary conditions. If it does, $P_{n_1, n_2, \ldots, n_s}(B)$ is unity; if it does not, zero. Hence we get the simpler expression

$$P_B(n_1, n_2, \ldots, n_s) = \frac{P(n_1, n_2, \ldots, n_s)}{\Sigma\, P(n_1, n_2, \ldots, n)_s},$$

the summation this time being taken over only those sets which satisfy the conditions B.

Now, no matter what set of values of the n's we may choose, so long as it is an admissible set, the denominator of this fraction is always the same. It is, in fact, a constant of such a nature that, if the *unconditional* probability of any admissible set of values is divided by it, the result is the *conditional* probability of that same set. It will therefore be simpler for us to write simply

$$P_B(n_1, n_2, \ldots, n_s) = K\, P(n_1, n_2, \ldots, n_s),$$

since we can always determine the value of the constant K when we need it by means of the fact that the sum of such terms, taken over every *admissible* set of n's, must be unity.

This effects a very great simplification in our problem, for it means that, except for a multiplicative constant, our probabilities have exactly the same formal definition no matter whether they are subject to auxiliary conditions or not. Because of this we may as well write our approximation (134) in the form

$$P(n_1, n_2, \ldots, n_s) = K\, e^{-\frac{1}{2}\Sigma x_i^2}, \qquad (136)$$

which may then serve our purpose *in every case* provided the K is given a suitable value; and by a suitable value we shall mean *in every case* that value which makes the sum of the probabilities unity.

§ 102. *The Measure of the Goodness of Fit*, $P(>\chi^2)$

In the discussion of Tables XXI and XXII we are actually interested in just 13 different variables, which we have denoted by the symbols $n_0, n_1, \ldots, n_{12}$; but it will be wise for us to think for a moment in terms of only three in order that we may be able to visualize certain geometrical ideas. Suppose, then, that we had only the three variables n_1, n_2, n_3 so that (136) reduced simply to

$$P(n_1, n_2, n_3) = K e^{-\frac{1}{2}(x_1^2+x_2^2+x_3^2)}.$$

If two sets of values of these n's are equally likely, the exponential factors in their probabilities must be equal. Conversely, any two sets of n's are equally likely for which the relation

$$x_1^2 + x_2^2 + x_3^2 = r^2 \qquad (137)$$

is satisfied. Written explicitly in terms of the n's this relation becomes

$$\frac{(n_1 - mp_1)^2}{mp_1} + \frac{(n_2 - mp_2)^2}{mp_2} + \frac{(n_3 - mp_3)^2}{mp_3} = r^2. \qquad (138)$$

If we were to represent sets of n's by points in three-dimensional space, all those points which lay upon the ellipsoid defined by (138) would have the same probability $P = K e^{-r^2/2}$, and we see at once that the smaller the value of r^2 is the *smaller* the ellipsoid and the *larger* the probability.

Corresponding to different values of r there are, therefore, a set of ellipsoids defined by the equation (138), all having the same center (mp_1, mp_2, mp_3), and no two intersecting. These ellipsoids form a sort of nest, one within another, of such a nature that the probability P decreases progressively as we go outward from their common center.

Suppose, now, that an experiment has given us a particular set of values of the three n's, and that we have computed the sum of the squares of the three x's which correspond to these values and found it to be χ^2. Suppose, moreover, that we have made estimates of the probabilities which underlay the experiment and are attempting to check the reasonableness of these estimates in accordance with the criterion laid down in § 97. What we must do is to add together the probabilities of all admissible sets of values which are less likely to occur than the experimental one. We know, however, that the points which correspond to these sets all lie outside the ellipsoid for which $r = \chi$. Hence to check the plausibility of our estimates we need only add together the probabilities corresponding to all those admissible points for which $r > \chi$.

As a method of carrying this out, we shall replace the summation over discrete points by an integration over continuous space, just as in § 93 we replaced a summation with respect to a single variable by an integral. To do this, however, we must first multiply (136) by the Jacobian of the transformation from the n's to the x's in order to obtain an expression for $P(x_1, x_2, \ldots, x_s)$. By actual computation from (135) this Jacobian is found to be $(\sqrt{m})^{s-1}\sqrt{p_1 p_2 \ldots p_s}$. As this is a constant so far as the x's are concerned its product by the K which is already present in (136) is just another constant, and we need no new symbol to represent it; for we must eventually compute the right value of K by making the sum of the admissible probabilities unity, and any multipliers which we may drop now will automatically reappear in this process. Hence we write

$$P(x_1, x_2, \ldots, x_s) = K e^{\frac{-r^2}{2}}, \qquad (139)$$

where r^2 is written briefly for the sum of the squares of x's.

Returning to our three dimensional case, we notice that what was the ellipsoid (138) in the space in which the n's were represented becomes the sphere (137) in x-space, with its center at the origin and its radius equal to r. This, indeed, is just the significance of the x's which we introduced into

(133) merely for simplicity of expression: they are deviations measured in such units that equal "vector deviations" are equally likely, no matter what their "directions."

Our integration, therefore, extends over all those admissible values which lie outside a sphere of radius χ. But before we can carry it out, we must know just what regions contain these admissible values. To this end we return to the two equations (128) and (130) and notice that they are both of the form

$$\Sigma\, a_j n_j = m \Sigma\, a_j p_j. \qquad (140)$$

In (130) the a's are all unity; in (128) $a_j = j$. Later on, in our discussion of the methods of obtaining empirical equations to represent statistical data, we shall find that all our estimates are arrived at through the use of equations of this form. We therefore restrict our discussion to this type of auxiliary condition, which becomes, in terms of the x's,

$$\Sigma\, b_j x_j = 0,$$

where the b's are constants related to the a's by the rule $b_j = a_j \sqrt{m p_j}$.

In our three-dimensional case, any one of these equations is of the form

$$b_1 x_1 + b_2 x_2 + b_3 x_3 = 0,$$

and represents a plane passing through the origin of coordinates. If there is only one such condition, then, the admissible points must all lie upon such a plane, and our integral is not a volume integral of (139) outside a certain sphere, but a surface integral outside a certain circle. If there are two such conditions, the admissible points are those which lie in *both* planes, and therefore upon the line in which they intersect. In that case our integral reduces to the integral of (139) over those portions of a line further removed from the origin than a certain experimentally determined point.

Now exactly this same situation exists in general. A single condition upon our variables (there is never less than one, because the condition (130) must always be satisfied)

reduces us from a space of s dimensions to one of $s - 1$ dimensions, and requires that we integrate (139) over all those portions of this space which are further from the origin than a certain predetermined amount χ. And q conditions reduce us to a space of $s - q$ dimensions, again requiring an integration over all that region that lies outside a hypersphere of radius χ. So far as computation is concerned, therefore, we are interested only in the integral of $K e^{-r^2/2}$ in a space of $s' = s - q$ dimensions.

We can get some clue to how next to proceed by considering what we would do in the case of a space of 1, 2 or 3 dimensions. In one dimension we would of course have immediately

$$2K \int_{\chi}^{\infty} e^{\frac{-r^2}{2}} dr.$$

In two dimensions we would take advantage of the fact that the integrand depends only on r and write our integral in the form

$$2\pi K \int_{\chi}^{\infty} e^{\frac{-r^2}{2}} r\, dr,$$

while in three dimensions we would write it as

$$4\pi K \int_{\chi}^{\infty} e^{\frac{-r^2}{2}} r^2\, dr.$$

It appears that in a space of s' dimensions the integrand would be of the form $e^{-r^2/2} r^{s'-1}\, dr$ multiplied by some constant the law of formation of which is not apparent.[1] But we have no occasion to know its exact value, since it can be amalgamated with the unknown constant K whose value we must eventually determine anyway. We thus arrive at the result

$$P_{r'}(> \chi^2) = K \int_{\chi}^{\infty} e^{\frac{-r^2}{2}} r^{s'-1}\, dr. \qquad (141)$$

[1] The constant is actually $(\sqrt{2\pi})^{s'}/(\frac{1}{2}s')!$, as we could readily show if we cared to introduce hyperspherical coordinates. For our purposes we need not introduce this form of coordinate system, which is probably unfamiliar to the student.

We are now ready to determine the value of K. We have already said that the K must have such a value that the sum of *all* admissible probabilities is unity. But that sum is obviously just the integral (141) taken over every value of r from zero upward. This gives us

$$1 = K\int_0^\infty e^{\frac{-r^2}{2}} r^{s'-1}\,dr,$$

which readily reduces to the form

$$1 = K\int_0^\infty e^{-u}(2u)^{\frac{1}{2}s'-1}\,du$$

if we make the substitution $r^2/2 = u$. Comparing this with (6), however, we arrive at once at the value

$$K = \frac{1}{2^{\frac{1}{2}s'-1}(\frac{1}{2}s'-1)!}.$$

Substituting this in (141) we have as our final result

$$P_{s'}(>\chi^2) = \frac{1}{2^{\frac{1}{2}s'-1}(\frac{1}{2}s'-1)!}\int_\chi^\infty e^{\frac{-r^2}{2}} r^{s'-1}\,dr. \qquad (142)$$

This formula, like many others of frequent use, has been reduced to the form of tables, one of which is given in Appendix VIII. The values of P are given in the headings of the columns, and the values of χ^2 in the body of the table. In order to find the probability of any estimate, therefore, it is only necessary to determine s' and χ^2, whereupon the approximate value of P can be taken at once from the table.

§ 103. *The Solution of Examples 50 and 51*

We are now prepared to renew our discussion of Weldon's dice data, with the direct purpose of determining whether the fact that his results were classified as indicated in Table XX affects our conclusion that the dice were probably biased.

Our theoretical considerations have taught us that the clue to this question lies in the magnitude of the quantity which

we have called χ^2. We must therefore consider how it is to be computed.

We notice that by definition x_j^2, which is usually called the "divergence" of n_j, is

$$x_j^2 = (1 - p_j)\, y_j^2 = \frac{\delta_j^2}{\epsilon(n_j)}.$$

To get χ^2, it is therefore only necessary to divide the square of each deviation by the expectation from which the deviation was measured, and then add the results.

This has been done in connection with Tables XXI and XXII, with the results shown in the last columns. The only point about the table which is likely to occasion any surprise is the grouping of the last two entries in computing the divergence, the two numbers being added and used as if they were one. The reason for this is to be found in the theory underlying our derivation of the χ^2 test: in which we concluded that the substitution of Stirling's Formula for the factorial was not justifiable when the number n_j was too small. How we divide the data up, however, is a matter entirely for our own judgment, and we can always place the divisions in such a way that a fair number of observations will fall in each.

Upon the assumption that the dice are true (Example 50) we obtain $\chi^2 = 40.75$. Due to grouping the last two entries in Table XXI there are just 12 variables; and as they must satisfy (130) only 11 of them can be independent. Hence in entering the table of χ^2 in Appendix VIII to find the probability of so large a deviation, we must make use of the set of values corresponding to $s' = s - q = 11$. The number 40.75 is well off the right-hand side of the table; hence the probability of such an occurrence with true dice is very much less than 0.01. From a more extensive table it would be found to be 0.000,03. Looked at as a whole, as we did in § 100, Weldon's result appeared to be even less likely, but from either standpoint it appears highly probable that the dice were inaccurate.

Passing now to Example 51 we find χ^2 to be 12.68. But in computing *this* table, we have made use of *two* linear relations

among the n_j's: (128) and (130). Hence this time in entering Appendix VIII for $P(> \chi^2)$ we must reduce the number of variables by 2, and use the row marked $s' = 10$. We find $P = 0.25$. That is, with identical biased dice for each of which the probability of throwing a 5 or 6 was 0.3377, a divergence as great as that observed could be expected about once in four times.

I must emphasize again that these figures, taken only by themselves, *decidedly do not say that the dice were probably biased.* They only force one of two conclusions upon us:

1. *Either* the dice are biased,

2. *Or* a very unexpected thing has happened.

It is only because we feel that conclusion (1) is a plausible one — as of course we do in this case — that we choose it in preference to the less plausible (2).

§ 104. *Résumé of the Test of Goodness of Fit*

The discussion of the test of goodness of fit has occupied so many pages, and has been punctuated by so much mathematics of a kind with which the student is probably none too familiar, that it seems wise before leaving it to sketch hastily the main facts to which we have been led.

We begin with the not unreasonable postulate that before an experiment is performed, there is a certain probability of it giving a specified result. We may not know what that probability is — if we did there would usually be no need for a test of goodness of fit; but at any rate such a probability exists. Also there is more than one possible result. Hence under the conditions of the experiment there are a group of probabilities, one for each possible result, which may be represented as a distribution function — or better, might be so represented if we knew what their values are.

The difficulty which confronts us is exactly that we do not know what they are, but are trying by experiment to determine them. Let us, however, ignoring this difficulty for a moment, *assume* that they are represented by the distribution

function of Fig. 35, and seek to learn how unusual the result of our experiment is.

The nature of the experiment may be such that only a discrete set of events may occur — as was the case with the dice problem which we have been considering — or it may concern itself with such a variable as length, which can take any value whatever. No matter which of these is true, we may divide the entire range of variation into intervals (or classes) such as those shown in Fig. 35. If the variable can take only a discrete set of values, it may often be natural to make each of these values a distinct class; we did this in the case of the dice experiment, except in the case of the values 11 and 12 which occurred so seldom that we classed them together. But this is not necessary. We can classify our data pretty much as we see fit. The classes need not even contain equal ranges of the variable, as an indication of which we have made them obviously unequal in the figure.

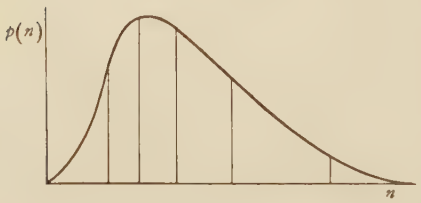


FIG. 35.

To each of these classes corresponds a very definite probability. Hence we can compute the probability of m events partitioning themselves among the classes in such a way that there are exactly n_j in the jth class. When we do so we obtain the Multinomial Law (129), which is too complicated for purposes of computation. We find, however, that it is quite accurately represented by the generalized Normal Law (139), provided none of the numbers n_j is too small. This generalized Normal Law has, as its single variable, the quantity r^2 formed from the sum of the "divergences" of the individual classes, the term "divergence" meaning the square of the deviation of n_j from its expectation, divided by the expectation.

So much for pure *a priori* reasoning. Usually, however, when we conduct an experiment we do not know the exact form of the distribution function pictured in Fig. 35. Instead,

we are often attempting to find out what it is. Having gotten our experimental results we attempt as best we can to infer what the distribution function is like and arrive at some estimates as to the probabilities associated with the various class intervals of Fig. 35. Next we naturally attempt to check up the reasonableness of our results; and in doing this are led to the idea of finding how probable it is that an experiment, conducted under our *assumed* conditions, would give a result that is at least as unusual as the experimental one.

To do this, we find the sum of the divergences of the experimental data, which we call χ^2, and then ask for the chance of an experiment giving a less likely (that is, a larger) value of r^2 than this. To get the answer to this question it is necessary to integrate (139) over all values of r which exceed χ, the result of which integration is contained in the formula (142), or for practical purposes in Appendix VIII. Having thus obtained $P(> \chi^2)$, if it is not too large we conclude that our assumed distribution function is a plausible one, *in the sense that the observations would not be miraculous if it were the correct one*, while if the probability is very small, we conclude that the experiment was probably conducted under conditions which differed materially from those assumed. However, the assurance we feel in these conclusions must be tempered by our judgment as to their inherent plausibility. We must not, for instance, accept a preposterous assumption as justified merely because we get a high value of P; or one which is almost certain to be true as being disproved by a low value of P: for a low value of P merely says that the result is unusual, not that it could not occur.

When we compute P we find that it has different values according to the number of classes into which we have divided our range of possibilities. Hence the table of P's is a double-entry table, arranged according to values of χ^2 and the "class-number" s'. Moreover, we find that this class-number is not the total number of classes s, but the number that, under the conditions of the problem, may be assigned values arbitrarily — that is, the number of *independent* classes. Since

the sum of all the n_j's must equal the total number of events, no more than $s - 1$ of them can ever be assigned arbitrarily; and if in addition we make use of our data to compute constants of our assumed formula, as in the case of Table XXII we used the data to find the appropriate value of p, we must reduce the class-number by one for each such condition. This is the only difficulty in the *application* of the test for goodness of fit; its *justification*, as we have seen, is another matter.

In the sections which follow we give a few more illustrations of its use.

§ 105. *Some Instructive Illustrations; Some Telephone Data*

In Chapter X we shall find that there is reason to believe that the probability of just n pieces of telephone apparatus being in use at a given instant is often given by the Poisson Formula. One case, in particular, to which we should expect this formula to apply reasonably well is the type of automatic equipment known as the "sender." Hence as a second illustration we take the data contained in Table XXIII, which covers 3754 observations upon the number of senders busy in a panel type machine switching exchange.

The Poisson Law is determined solely by its expectation ϵ. In the case of our data the average number of busy senders is 10.44, and if we use this value as our estimate of ϵ the Poisson Formula gives the values shown in the third column under the heading "Expected Frequency." The deviations from expectation are listed in the fourth column; the divergences in the fifth, and their total, which is χ^2, at the bottom. The quantities are all large enough that there is no occasion to combine any of the classes, except in the case of the top two (0 and 1). Hence there are in all 22 classes. However, since we have determined our expectation from the data only 20 of these classes are independent, for the only frequencies with which we dare compare our experimental results are those which have a sum 3754 and an average 10.44. We therefore look up the entry 43.43 in that portion of Appendix

VIII which corresponds to the class number 20, and find that it is beyond the right-hand margin of the table — apparently at about $P = 0.005$.

TABLE XXIII

NUMBER OF BUSY SENDERS IN A TELEPHONE EXCHANGE

Number Busy	Observed Frequency	Expected Frequency	Deviation, δ	Divergence, x^2
0	0	0.11	+ 3.74	11.01
1	5	1.15		
2	14	5.98	+ 8.02	10.76
3	24	20.82	+ 3.18	.49
4	57	54.33	+ 2.67	.13
5	111	113.44	− 2.44	.05
6	197	197.38	− 0.38	.00
7	278	294.38	−16.38	.91
8	378	384.16	− 6.16	.10
9	418	445.63	−27.63	1.71
10	461	465.24	− 4.24	.03
11	433	441.56	− 8.56	.17
12	413	384.15	+28.85	2.17
13	358	308.50	+49.50	7.94
14	219	230.05	−11.05	.53
15	145	160.11	−15.11	1.43
16	109	104.47	+ 4.53	.20
17	57	64.16	− 7.16	.80
18	43	37.21	+ 5.79	.90
19	16	20.45	− 4.45	.97
20	7	10.67	− 3.67	1.26
21	8	5.31	+ 3.69	1.36
22	3	4.51	− 1.51	.51
		3753.77	+ 0.23	$x^2 = 43.43$

The fit is none too good, and we would probably reject this solution as unsuitable, were it not for the theoretical basis which underlies the choice of the Poisson Formula, and which we cannot entirely overlook when considering the significance of such a low number.

§ 106. *Some Instructive Illustrations; Chips Drawn from a Normal Universe*

Our next illustration will be one in which the conditions of the experiment are controlled artificially, so that the data can be said in advance to follow a certain law.

EXAMPLE 52.—*By reference to Appendix V it can be found that if a variable is governed by the Normal Law it has a chance 0.1974 of lying between − 0.25 and + 0.25; a chance 0.1747 of lying between − 0.25 and − 0.75, and so on, as shown in the third column of Table XXIV. An experiment was performed by marking 197 chips with the number 0; 175 with each of the numbers − 0.5 and + 0.5; and so on, the number carrying each marking being the same as the first three digits of the probability corresponding to the interval in which this number lay. These chips were then placed in a box, thoroughly mixed, and one drawn out. After its marking had been noted it was replaced, the contents of the box again mixed, and another drawing made. The results of 1000 such drawings were distributed as shown in the second column of Table XXIV.*

What is the probability that another similar experiment would deviate from expectation as much as this one did?

The entire computation is shown in Table XXIV. The number expected is just the theoretical probability multiplied by 1000; hence the only condition imposed upon our theoretical variables is that their sum shall equal the number of drawings. As there are 13 classes in the table, this leaves 12 independent, and we enter the table of χ^2 with the class number $s' = 12$. The answer to the question asked in the example is therefore 0.74. This is a very high probability — our drawings were nearer to the Normal Law than we could reasonably have expected.

§ 107. *The Determination of a Suitable Distribution Function when no Theoretical Formula is Known*

We now have a method for testing how well our assumed distribution fits the data, but so far we have applied it only to data which we had reason to believe followed some one of our well-known laws. In many cases we have no such information in advance and are forced to make a purely empirical choice of the curve. Our next problem is to build up a systematic method of attacking this problem.

To begin with, we remember that moments and expectations are correlated ideas, the moments being derived directly from data and the expectations being the analogous quantities as computed from the distribution functions. An actual experiment is not likely to give us moments that are *exactly* equal to their corresponding expectations, but we have seen in Chapter VII that if the experiment is extensive the chance

TABLE XXIV

AN ARTIFICIALLY CONTROLLED EXPERIMENT

Marking	Number Observed	Theoretical Probability	Number Expected	Deviation	Divergence
−3.0	5	0.0024	2.4	+ 2.6	2.82
−2.5	9	0.0093	9.3	− 0.3	0.01
−2.0	36	0.0278	27.8	+ 8.2	2.42
−1.5	55	0.0656	65.6	−10.6	1.71
−1.0	123	0.1210	121.0	+ 2.0	0.03
−0.5	165	0.1747	174.7	− 9.7	0.54
0.0	203	0.1974	197.4	+ 5.6	0.16
+0.5	172	0.1747	174.7	− 2.7	0.04
+1.0	123	0.1210	121.0	+ 2.0	0.03
+1.5	68	0.0656	65.6	+ 2.4	0.09
+2.0	31	0.0278	27.8	+ 3.2	0.37
+2.5	8	0.0093	9.3	− 1.3	0.18
+3.0	2	0.0024	2.4	− 0.4	0.07
	1000	0.9990	999.0	+ 1.0	$\chi^2 = 8.47$ $n = 12$ $P = 0.74$

of any great disagreement between them is small. It seems natural, then, when we know nothing about the nature of the distribution in advance, to assume that the distribution function has expectations equal to the moments of the data. The assumption is false, of course, but it is probably as near the truth as any we can make.

Our process of fitting data will then have, as its first element, the location of some distribution type which, by suitable choice of the arbitrary constants in its equation, may be made to have the desired set of expectations. In searching

for such a suitable type we shall find the measures of asymmetry and flatness of considerable assistance.

If a distribution function is symmetrical all the expectations of odd order are zero, for negative and positive deviations have like probabilities and therefore destroy one another. It therefore seems natural to measure asymmetry by means of these odd expectations. However, the first expectation cannot be used, for it is zero by definition; hence it is customary to make use of the third. But if it is to measure a property of the *shape* of the curve, and not be affected by the *scale* to which it is drawn, it is necessary to specify a standard scale of measure, which is accomplished by requiring that the deviations be measured in terms of their own standard deviation as a unit. This gives us the following definition:

The third expectation of the deviations of a variable n, expressed in terms of its own standard deviation as a unit, is the asymmetry[1] *of the distribution function which n obeys. It is denoted by the symbol* $\sqrt{\beta_1}$.

In addition to this measure of asymmetry, however, we need also a measure of flatness. To illustrate what we mean by this term, we may refer to Fig. 36, both curves of which are symmetrical and have the same standard deviation, though they differ from one another to an important degree. We phrase this difference by saying that B is "flatter" than A.

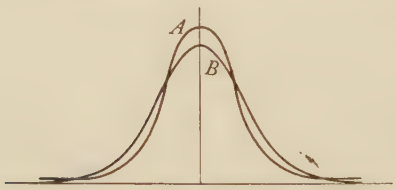


Fig. 36.

The portions of area outside A and inside B are just equal to the portions inside A and outside B, otherwise the curves would not bound equal areas; and they are so shaped that their second expectations are equal. Otherwise the two would not have the same standard deviation. But the fourth expectation of A is obviously bigger than the fourth expectation of B, due

[1] Also called "skewness."

to the fact that the larger tail of A plays an increasingly important part the higher the order of the expectation. Hence the fourth expectation affords a method of measuring flatness. The exact definition is:

The fourth expectation of the deviations of a variable n, measured in terms of its own standard deviation, is the flatness[1] *of the distribution function which n obeys. It is denoted by the symbol β_2.*

We must observe at once, however, that in computing asymmetry and flatness we need not bother to reduce all our data to the units specified in the definitions. In fact, the third expectation is by definition the sum of quantities each of which is the product of a probability by the cube of a deviation. The probability being a pure number the third expectation varies as the inverse cube of the unit in terms of which the deviations are measured. That is, we may compute our third expectation to any scale we please, and then divide it by σ^3 in order to reduce it to the units demanded by the definition of asymmetry. Similarly, the flatness will ordinarily be obtained from the fourth expectation divided by σ^4.

Quite similar concepts can be defined in the case of sets of experimental data, except that we deal then with *moments* instead of *expectations*. We may indicate them sufficiently well by means of the following parallel résumé:

Properties of Distribution Functions	*Properties of Experimental Data*
The first expectation of a variable n is defined by (83).	The first moment of a set of data is their *average*, as defined by (82).
Deviations δ are measured from this expectation.	Deviations d are measured from this average.
The first expectation of δ is therefore $\epsilon_1 = 0$. [See (96).]	The first moment of the set of d's is therefore $\mu_1 = 0$. [See (93).]
The square root of the second expectation of δ is called the "standard deviation." That is, $\sigma = \sqrt{\epsilon_2}$.	The square root of the second moment of the deviation is called the standard deviation. That is, $\sigma = \sqrt{\mu_2}$.

[1] Also called "kurtosis."

Properties of Distribution Functions	*Properties of Experimental Data*
The asymmetry is the third expectation, provided the deviations are measured in terms of σ as a unit. In general the formula is $\sqrt{\beta_1} = \epsilon_3/\sigma^3$.	The asymmetry is the third moment, provided the deviations are measured in terms of σ as a unit. In general the formula is $\sqrt{\beta_1} = \mu_3/\sigma^3$.
The flatness is the fourth expectation, provided the deviations are measured in terms of σ as a unit. If not, the formula for it is $\beta_2 = \epsilon_4/\sigma^4$.	The flatness is the fourth moment, provided the deviations are measured in terms of σ as a unit. If not, the formula for it is $\beta_2 = \mu_4/\sigma^4$.

In addition to these characteristics there are a number of others, with less precise geometrical significance, which have been found of service in sorting out the type of curve which is most suitable for a particular set of data. The only one which need claim our attention is the combination of quantities

$$J = 3\,\beta_1 - 2\,\beta_2 + 6,$$

to which we shall give the name "Type Criterion." For some types of distribution curves this type criterion is always positive, for some negative, and for still others zero. Naturally, then, it is a helpful thing to have.

In Appendix XI we have listed the equations of the Normal, Poisson and Binomial Laws, the Gram–Charlier Series, and the four Pearson Types which we have discussed; and underneath them have written out the general formulæ for the quantities $\epsilon_1(n)$, ϵ_2, ϵ_3, ϵ_4, σ, $\sqrt{\beta_1}$ and β_2, and have also indicated the sign of the type criterion for each. Appendix XI, then, is of the nature of a compendium of all the basic information which underlies the fitting of curves to data.

But we can do more than this. We can find, by suitable investigations, that the only ones of these curves which can have, at the same time, $\beta_1 = 0$ and $J < 0$ are the Pearson Type IV and the Gram–Charlier Series. Similarly the only ones for which both β_1 and J can be zero are the Normal Law and the Gram–Charlier Series. In this way we can consider in turn every possible combination of values of the

quantities β_1 and J. When the results are arranged schematically they lead to the table given in Appendix X.[1]

Suppose, then, that we have a set of data to which we wish to fit a suitable distribution curve. From it we may easily obtain its various moments, and thus derive the five quantities

$\bar{n}$, average,
σ, standard deviation,
$\sqrt{\beta_1}$, asymmetry,
β_2, flatness,
J, type criterion.

Then by entering Appendix X with the values of β_1 and J we may sort out certain types which seem appropriate for our purpose. Finally, we may so determine the arbitrary constants in our chosen equation that the expectations of the distribution function are equal to the moments of the data.[2]

Much of the labor of this final step may be carried out once for all by algebraic means, leading to formulæ in which it is only necessary to substitute the values of the quantities $\bar{n}$, σ, β_1 and β_2 in order to determine the constants of the equation directly.

Appendix XI also contains the equations which are needed for this purpose, listed with the title " Equations for Determining Constants."

We have then in Appendices X and XI all the information which is needed for the determination of our distribution

---

[1] It must be remembered that β_1 is the square of the asymmetry and therefore can never be negative.

[2] We should not fail to observe that this process leads us to equations of exactly the form (140). For the ith expectation of the variable (which was called j in § 102) is

$$\epsilon_i(j) = \Sigma j^i p_j;$$

and the ith moment of the observed values is

$$\mu_i(j) = \Sigma j^i n_j.$$

Equating the two leads at once to (140), the coefficients being $a_j = j^i$.

curve, provided, of course, that any one of the eight types which we have been discussing is suitable for the purpose. From this point on we can best discuss the problem of "curve fitting" by the consideration of several examples.

§ 108. *Some Instructive Illustrations; A Reconsideration of Weldon's Dice Data*

As a first illustration, we consider again the data presented in Table XX. If we choose to forget its origin, we may ask ourselves to obtain an appropriate distribution law for it. We first determine the various moments of n as shown in Table XXV. Then by the use of the formulæ [1]

$$\mu_2(d) = \mu_2(n) - \overline{n}^2,$$

$$\mu_3(d) = \mu_3(n) - 3\mu_2(d)\,\overline{n} - \overline{n}^3,$$

$$\mu_4(d) = \mu_4(n) - 4\mu_3(d)\,\overline{n} - 6\mu_2(d)\,\overline{n}^2 - \overline{n}^4,$$

we compute the various moments of the deviations, and then in turn the standard deviation, asymmetry and flatness.

As a result of all these computations it turns out that both β_1 and J are greater than zero. Entering Appendix X with this information we find that either the Pearson Type I, the Binomial or the Poisson Formula (as well, of course, as the Gram-Charlier Series which can be used for any type of data) satisfies the conditions of the problem. Let us, then, attempt to treat the data by means of all of these methods except the Pearson one (for which the computations are difficult), and see which gives the most satisfactory fit. Our knowledge of the origin of the data would lead us to expect that the Binomial Law would be better than any of the others, but we are laying aside this part of our knowledge regarding the data and treating it as if it had no theoretical significance.

---

[1] The first of these is identical with (94). The rest are obtained in just the same way as (94).

TABLE XXV

THE MOMENTS AND RELATED STATISTICS OF WELDON'S DICE DATA

n	Observed Frequency f	nf	n^2f	n^3f	n^4f
0	185	0	0	0	0
1	1,149	1,149	1,149	1,149	1,149
2	3,265	6,530	13,060	26,120	52,240
3	5,475	16,425	49,275	147,825	443,475
4	6,114	24,456	97,824	391,296	1,565,184
5	5,194	25,970	129,850	649,250	3,246,250
6	3,067	18,402	110,412	662,472	3,974,832
7	1,331	9,317	65,219	456,533	3,195,731
8	403	3,224	25,792	206,336	1,650,688
9	105	945	8,505	76,545	688,905
10	14	140	1,400	14,000	140,000
11	4	44	484	5,324	58,564
12	0	0	0	0	0
	26,306	106,599	502,970	2,636,850	15,017,018

$$\bar{n} = 4.0522694$$

$$\mu_2(n) = 19.119972$$

$$\mu_3(n) = 100.23759$$

$$\mu_4(n) = 570.85904$$

$$\mu_2(d) = \mu_2(n) - \bar{n}^2 = 2.699085$$

$$\mu_3(d) = \mu_3(n) - 3\mu_2(d)\,\bar{n} - \bar{n}^3 = 0.88347$$

$$\mu_4(d) = \mu_4(n) - 4\mu_3(d)\,\bar{n} - 6\mu_2(d)\,\bar{n}^2 - \bar{n}^4 = 20.9651$$

$$\sigma = \sqrt{\mu_2(d)} = 1.642889$$

$$\beta_1 = \frac{[\mu_3(d)]^2}{[\mu_2(d)]^3} = 0.039650$$

$$\sqrt{\beta_1} = 0.19912$$

$$\beta_2 = \frac{\mu_4(d)}{[\mu_2(d)]^2} = 2.87782$$

$$J = 0.36331$$

Let us first consider the Binomial case. Referring to Appendix XI, we find that the constants of the formula are to be determined from the relations

$$p = 1 - \frac{\sigma^2}{\bar{n}},$$

$$m = \frac{\bar{n}}{p}.$$

Thus we derive $p = 0.3339325$, and $m = 12.13500$. Using these values, we obtain for our empirical distribution function the formula

$$p(n) = C_n^{12.135}\,(0.339325)^n\,(0.660675)^{12.135-n}.$$

From this formula the values shown in the third column of Table XXVI were obtained.[1]

To get the expected frequency it is only necessary to multiply each probability by 26,306. Finally, the divergence is computed in the usual way.

The resultant criterion as to goodness of fit is $\chi^2 = 11.59$, which is really smaller than that obtained in § 100. But we must not be misled by this fact. In the first place, the theoretical foundation underlying the Binomial Law strongly suggests that the value of m should be integral, and the use of the

[1] The process of computation was as follows: The value of $p(0) = (0.660675)^{12.135}$ was first found by the use of logarithms. Then it was noted that

$$\frac{p(n)}{p(n-1)} = \left(\frac{12.135}{n} - 1\right)\left(\frac{0.339325}{0.660675}\right).$$

With the aid of a computing machine the value of this ratio is easily obtained for each value of n. Call it r_n. Then the remaining probabilities are found from the formulæ

$$p(1) = r_1\,p(0),$$

$$p(2) = r_2\,p(1),$$

$$\cdots$$

$$p(n) = r_n\,p(n-1).$$

value 12.135 instead of 12 is itself open to suspicion, particularly so when we remember that there were really *twelve* dice. In the second place, in § 100 we required our formula to agree with the data in only *two* particulars: that $p = 0.3377$, and that the expected number of successes should equal the number

TABLE XXVI

An Empirical Binomial Formula for Weldon's Dice Data

n	Observed Frequency	Empirical Binomial Law		
		Probability	Frequency	Divergence
0	185	0.007218	189.9	0.13
1	1,149	0.043911	1,155.1	0.03
2	3,265	0.122567	3,224.2	0.52
3	5,475	0.207594	5,461.0	0.04
4	6,114	0.237687	6,252.6	3.07
5	5,194	0.193880	5,100.2	1.73
6	3,067	0.115589	3,040.7	0.23
7	1,331	0.050790	1,336.1	0.02
8	403	0.016345	430.0	1.70
9	105	0.003765	99.0	0.36
10	14	0.000592	15.6	0.16
11	4	0.000058	1.5	3.60
12	0	0.000003	0.1	
	26,306	0.999999	26,306.0	$\chi^2 = 11.59$

observed. Here we are requiring in addition that they agree also as to standard deviation. Hence, when we enter Appendix VIII to find $P(> \chi^2)$ we must use in the present instance the row marked $s' = 9$, whereas before we could use $s' = 10$. The result is, that we now have $P(> \chi^2) = 0.24$, where before we had $P(> \chi^2) = 0.25$. So even on the purely formal side, if we use our χ^2 criterion correctly, we have no better result than before.

Turning to the case of the Poisson Formula, we find from Appendix XI that the only constant upon which the formula depends (namely, ϵ), is equal to $\bar{n}$. Using this value we

obtain [1] the probabilities shown in the third column of Table XXVII; and from them the frequencies shown in the next column, and the divergences in the column following that.

In this case we have made our formula agree with the data in only two respects: the total number of successes and the average number. Hence it is only necessary to reduce the

TABLE XXVII

AN ATTEMPT TO FIT WELDON'S DICE DATA WITH A POISSON LAW

n	Observed Frequency	Empirical Poisson Law		
		Probability	Frequency	Divergence
0	185	0.01738	457.2	162.1
1	1,149	0.07044	1,853.0	267.5
2	3,265	0.14272	3,754.4	63.8
3	5,475	0.19278	5,071.3	32.1
4	6,114	0.19530	5,137.6	185.6
5	5,194	0.15828	4,163.7	254.9
6	3,067	0.10690	2,812.1	23.1
7	1,331	0.06188	1,627.8	54.1
8	403	0.03135	824.7	215.6
9	105	0.01411	371.2	190.9
10	14	0.00572	150.5	123.8
11	4	0.00211	55.5	74.5
12	0	0.00102	26.8	
	26,306	0.99999	26,305.8	$\chi^2 = 1648.0$

[1] The computations were carried out as follows: The formula is

$$p(n) = \frac{(4.0522694)^n e^{-4.0522694}}{n!},$$

and reduces to $e^{-4.0522694}$ when n is zero. This value may be found by the use of logarithms. Then it is observed that

$$\frac{p(n)}{p(n-1)} = \frac{4.0522694}{n},$$

the values of which are easily written down at once. Let us denote them by r_n. Then the p's are found by successive multiplications by these quantities r_n, just as in the case of the computations which led to Table XXVI.

In this case, however, the Poisson Law gives appreciable values for $p(n)$ when n exceeds 12. We have therefore interpreted the entry $n = 12$ as being equivalent to $n \geq 12$, the number 0.00102 which stands in the third column of the table being actually the chance of "twelve or more" successes as given by the Poisson Law.

number of classes by two in entering Appendix IX, thus using the marginal number $n = 10$. But the value $\chi^2 = 1648$ is so very great that P is extremely small. The fit in this case is very bad, and we are thoroughly justified in the presumption that the data did not obey the Poisson Law.

Finally, if we make use of a Gram–Charlier Series, we have from Appendix XI the relations:

TABLE XXVIII

An Attempt to Fit Weldon's Dice Data with a Gram–Charlier Series

n	Observed Frequency	Empirical Gram–Charlier Law		
		Probability	Frequency	Divergence
0	185	0.00989	260.2	21.73
1	1,149	0.04521	1,189.3	1.37
2	3,265	0.12076	3,176.7	2.45
3	5,475	0.20547	5,405.1	0.90
4	6,114	0.23630	6,216.1	1.68
5	5,194	0.19257	5,065.7	3.25
6	3,067	0.11545	3,037.0	0.30
7	1,331	0.05209	1,370.3	1.13
8	403	0.01732	455.6	6.07
9	105	0.00415	109.2	0.16
10	14	0.00069	18.2	0.97
11	4	0.00009	2.4	1.07
12	0			
	26,306	0.99999	26,305.8	$\chi^2 = 41.08$

$\alpha = \bar{n} = 4.0522694,$

$\sigma = 1.642889,$

$A_1 = 0,$

$A_2 = 0,$

$$A_3 = -\frac{\sqrt{\beta_1}}{6} = -0.033187,$$

$$A_4 = \frac{\beta_2 - 3}{24} = -0.00509083.$$

Hence our empirical series takes the form

$$p^*(n) = \frac{1}{1.642889}(\phi - 0.033187\,\phi''' - 0.00509083\,\phi^{iv}), \quad (143)$$

the argument of the ϕ being, in every case, $\dfrac{n - 4.0522694}{1.642889}$.

By the use of this series [1] the probabilities shown in the third column of Table XXVIII were derived; and from them the expected frequencies and divergences shown in the next two columns. The value of χ^2 is found to be 41.08. In finding the probability of so large a value we must give due regard to the fact that we have forced our empirical law to conform to the data in five respects: in the total number of successes observed, and in the first four moments of n. Hence

---

[1] The Gram–Charlier Series, unlike the other two laws of which we have made use, allows its argument to vary continuously. The same is true of the various Pearson Types. Hence the remarks which follow apply to them also.

When dealing with experimental data which assigns a certain observed frequency to each of a group of discrete values of n, either because n is incapable of taking intermediate values, or because the observations have been arbitrarily classified in this way, we are forced to divide our theoretical distribution curve into a number of segments which correspond as well as may be with the classes into which the data has been grouped. In the present instance the experimental data is given for integral values of n ranging from 0 to 12: so we make an equivalent set of divisions in the case of the Gram–Charlier Series. To correspond to $n = 1$ we take the range $0.5 < n < 1.5$; to correspond to $n = 2$ the range $1.5 < n < 2.5$, and so on. There is no difficulty about this choice, which seems perfectly natural in view of the considerations set forth in § 93. The two end values, $n = 0$ and $n = 12$, however, are not so clear-cut, for the series would allow values of n as small or as large as we please. If we were to use ranges of unit width about these values as the centers, we would exclude the tails of our distribution function entirely, which seems not to be allowable since use has been made of them in evaluating the formulæ contained in Appendix XI. We choose what appears to be the only alternative: we regard the end ranges as extending from $-\infty$ to 0.5 and from 11.5 to ∞, respectively.

The probability corresponding to each value of n is now represented by a segment of area under the distribution curve, somewhat similar to those in Fig. 35. Obviously the function is varying too rapidly to permit us to use the ordinate in the middle of the segment as a satisfactory approximation to the area. We are forced, then, to compute the probabilities which appear in the third column of Table XXVIII by actual integration. We find, however, that the integral of (143) is

$$P(n) = \int p^*(n)\, dn = \phi_{-1} - 0.033187\,\phi'' - 0.0050908\,\phi''''.$$

By this we mean, of course, the indefinite integral. In terms of it, the areas of the segments, which are our desired probabilities, are expressed as follows:

$$p(0) = P(0.5) - P(-\infty), \qquad p(1) = P(1.5) - P(0.5),$$

$$p(2) = P(2.5) - P(1.5), \qquad \ldots, \qquad p(12) = P(\infty) - P(11.5).$$

only 7 of the 12 classes into which our data is divided can be regarded as independent. We find that $P(> \chi^2)$ is very small indeed. (From a larger table than that of Appendix VIII it can be found to be approximately 0.000001.)

What, now, can we learn from all this computation? We have seen that our criteria β_1 and J automatically eliminated from consideration all but four of our numerous empirical laws. We have also seen how the computations may be carried out for three of these four, and how we may test the excellence of the fit obtained from each of them. Finally, we have found that, even had we had no advance knowledge of the true law, our process would have picked it out for us as being by very long odds more likely than either of the others — always remembering, of course, that such a conclusion is only justified

---

In further explanation of this process of computation we present in Table XXIX the computations exactly as they were carried out.

TABLE XXIX

Computation of the Probabilities in Table XXVIII

x_j	$x_j - \alpha$	$\frac{(x_j - \alpha)}{\sigma}$	ϕ_{-1}	ϕ''	ϕ'''
$-\infty$	$-\infty$	$-\infty$	0.000000	0.000000	0.000000
0.5	−3.552269	−2.162209	0.015302	0.141575	0.139522
1.5	−2.552269	−1.553525	0.060151	0.168692	−0.108767
2.5	−1.552269	−0.944841	0.172373	−0.027393	−0.508300
3.5	−0.552269	−0.336157	0.368378	−0.334412	−0.365878
4.5	0.447731	0.272526	0.607390	−0.355840	0.306482
5.5	1.447731	0.881210	0.810897	−0.060467	0.530140
6.5	2.447731	1.489894	0.931874	0.160388	0.152848
7.5	3.447731	2.098578	0.982072	0.150168	−0.129979
8.5	4.447731	2.707262	0.996608	0.064676	−0.119762
9.5	5.447731	3.315945	0.999543	0.016334	−0.043325
10.5	6.447731	3.924629	0.999957	0.002599	−0.008782
11.5	7.447731	4.533313			
∞	∞	∞	1.000000	0.000000	0.000000

The first three columns are self-explanatory. The next three are values read from Appendix V. In the case of positive arguments they are read off directly, while

provided there are no reasons, not included in the mathematical computations, for hesitancy in accepting it.

§ 109. *Sheppard's Corrections to Moments Computed from Classified Data*

In our illustrations we have carefully avoided the use of any data which should logically be distributed continuously, and have confined our attention solely to such data as naturally falls into discrete classes. The reason for this has been to avoid a difficulty which presents itself in computing the moments of continuously distributed data, when that data has been artificially classified.

To see the nature of this difficulty let us consider the distribution curve shown in Fig. 37, and in particular the " class " between the values n_1 and n_2. This class ought, theoretically,

---

in the case of negative arguments use must be made of the fact that ϕ'' is an even function and ϕ''' an odd one, and also of the fact that $\phi_{-1}(-n) = 1 - \phi_{-1}(n)$.

TABLE XXIX.—*Continued*

n_j	$A_3\phi''$	$A_4\phi'''$	$P(n_j)$	$p(n)$	n
$-\infty$	0.000000	0.000000	0.000000		
				0.00989	0
0.5	−0.004698	−0.000710	0.009893		
				0.04521	1
1.5	−0.005598	0.000554	0.055106		
				0.12076	2
2.5	0.000909	0.000588	0.175870		
				0.20547	3
3.5	0.011098	0.001863	0.381338		
				0.23630	4
4.5	0.011809	−0.001560	0.617639		
				0.19257	5
5.5	0.002007	−0.002699	0.810205		
				0.11545	6
6.5	−0.005439	−0.000778	0.925657		
				0.05209	7
7.5	−0.004984	0.000662	0.977750		
				0.01732	8
8.5	−0.002146	0.000610	0.995071		
				0.00415	9
9.5	−0.000542	0.000221	0.999222		
				0.00069	10
10.5	−0.000086	0.000045	0.999915		
				000009	11
11.5					12
∞	0.000000	0.000000	1.000000		

These three columns having been written down, they were multiplied by the appropriate factors to give $A_3\phi''$ and $A_4\phi'''$, and then added to get $P(n_j)$. The computation of $p(n)$ then requires only the subtraction of each of these P's from the one which follows it.

to contribute to the ith *expectation* an amount $\int_{n_1}^{n_2} n^i\, p(n)\, dn$. If, however, the entire probability associated with the class is artificially assigned to the mid-point, we obtain a value $N^i \int_{n_1}^{n_2} p(n)\, dn$, which is not the same thing at all, unless the class interval is so exceedingly small that the quantities $p(n)$ and n^i are both substantially constant throughout it.

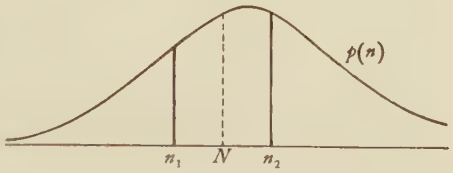


Fig. 37.

Now, what is true of these expectations in this regard is true also of the moments of any experimental data which we may classify in this fashion. Over and above the *accidental* differences between these moments and the theoretical expectations to which they correspond, there is a certain *regular* error due to the classification of the results: if by some odd turn of fortune the experimental data happened to be of just such a nature that its moments accurately corresponded with the expectations of its distribution law, the process of arranging the data in classes would destroy this agreement.

By studying the problem in the light of the Theory of Mechanical Quadratures, Sheppard was led to the conclusion that some of the error thus introduced could be eliminated by the use of the formulæ which follow. The unstarred quantities are the "classified" or "raw" moments; the starred ones are the "corrected" moments, and h is the "class interval" $n_2 - n_1$.

$$\left.\begin{aligned} \mu_1^*(n) &= \mu_1(n), \\ \mu_2^*(d) &= \mu_2(d) - \tfrac{1}{12} h^2, \\ \mu_3^*(d) &= \mu_3(d), \\ \mu_4^*(d) &= \mu_4(d) - \tfrac{1}{2} h^2 \mu_2(d) - \tfrac{1}{80} h^4. \end{aligned}\right\} \qquad (144)$$

Just what advantage these corrections possess is hard to say. It is not at all difficult to build up situations in which, instead of improving the moments, they make them less exact. But it must be admitted that these situations are always of a somewhat artificial sort, and that when a distribution of more usual aspect is considered the corrected moments are likely to be better than the "raw" ones. The consensus of opinion seems to be that they improve matters more often than not, and that they should be used. Certainly the error which they aim to eliminate is real enough; the only doubt concerns the possibility of correcting it by any other means than that of allowing continuous variation to the variable, which the limitations of instrumental measurement, for one thing, will not permit us to do.

We may take as an illustration the data presented in Table XXIV, in the derivation of which the chips were marked in classes in just the way in which an instrumental classification might arrange them, though the aim was to duplicate a Normal Universe. If the "raw" moments of this distribution of data are computed, the second and fourth (which are the only ones affected by Sheppard's corrections) are found to be 1.044 and 3.190, respectively, while the corrected moments are 1.023 and 3.062 instead. As the expectations of the Normal Law, to which the data was intended to correspond, are 1 and 3, the corrected moments are materially improved in this case.

§ 110. *The Distribution of Statistics*

Each of the quantities listed in Appendix XI is known as a statistic of its distribution law. "Statistic," then, is a general term meaning "moment," "expectation," "flatness," "asymmetry," "average," or indeed any other property computed from a distribution law, or from a set of data.

Suppose, now, that we have computed a "statistic" from a set of data — to make matters concrete, suppose we have computed the average of 50 observed numbers. What we thus get is not an absolute and invariable quantity which may be

reproduced whenever we wish by performing the experiment anew: indeed, another experiment would very likely give us a different result. All possible results are not equally likely, however. The values which we get for the average are governed by probability, in just the same way as any other quantity which is subject to accidental fluctuations. There is for it a certain distribution law, with its own standard deviation, asymmetry, and flatness. The same is true of any other "statistic" which we might name.

We may raise the question, therefore, as to how widely, and according to what law, such a statistic may be expected to vary. To go into this question in detail would involve us in an extended discussion of "precision of measurement," which is beyond the province of our text. We may mention in passing, however, that the four most important statistics — the average, the standard deviation, the asymmetry, and the flatness — have been shown to obey laws of distribution which are very nearly normal.[1] Hence, since the Normal Law is determined solely by its standard deviation, we may get a pretty fair estimate of the faith we are justified in putting in the observed values of any one of these statistics by knowing the standard deviation of the statistic in question. We give, in Table XXX and in Appendix IX the standard deviation of each of these statistics, when computed from a set of N data.

As an illustration of the use of these formulæ, we may compute the standard deviations of the statistics which we derived from Weldon's dice data, and which are presented in Table XXV. The formula for the standard deviation of the average, as taken from Table XXX is $\sigma(\bar{n}) = \sigma/\sqrt{N}$, which in the case of present numbers becomes

$$\sigma(\bar{n}) = \frac{1.642889}{\sqrt{26{,}306}} = 0.01013.$$

[1] They are *exactly* normal when derived from data which itself satisfies the Normal Law; and *approximately* normal in other cases.

Hence, in accordance with the usual practice in indicating the "precision" of our statistic, we should write

$$\bar{n} = 4.0523 \pm 0.0101,$$

as explained at the end of § 98.

It is interesting to note that this agrees with the value given in § 98 for the precision of p; for obviously, if the limits of $\bar{n}$ are those set forth above, the limits upon p should be one-twelfth as great. Thus we get,

$$p = 0.3377 \pm 0.0084,$$

as before.

TABLE XXX

STANDARD DEVIATIONS OF IMPORTANT STATISTICS

(N is the number of observations from which the statistic is computed)

Statistic			Standard Deviation of Statistic			
Name	Symbol	Symbol	General Formula	Special formula for		
				Normal Law	Poisson Law	Binomial Law
Average	$\bar{n}$	$\sigma(\bar{n})$	$\frac{\sigma}{\sqrt{N}}$	$\frac{\sigma}{\sqrt{N}}$	$\sqrt{\frac{\epsilon}{N}}$	$\sqrt{\frac{m p(1-p)}{N}}$
Standard Deviation	σ	$\sigma(\sigma)$	$\sqrt{\frac{\mu_4(d) - [\mu_2(d)]^2}{4N\mu_2(d)}}$	$\frac{\sigma}{\sqrt{2N}}$	$\sqrt{\frac{2\epsilon + 1}{4N}}$	$\sqrt{\frac{2(m-1)p(1-p) + \cdots}{4N}}$
Asymmetry (Skewness)	$\sqrt{\beta_1}$	$\sigma(\sqrt{\beta_1})$		$\sqrt{\frac{6}{N}}$		
Flatness (Kurtosis)	β_2	$\sigma(\beta_2)$		$\sqrt{\frac{24}{N}}$		

Next let us consider the standard deviation. By reference to Table XXX we find that $\sigma(\sigma)$ should be given by the formula

$$\sigma(\sigma) = \sqrt{\frac{\mu_4(d) - [\mu_2(d)]^2}{4N\mu_2(d)}}$$

which readily works out to be 0.00703. Hence it would be customary to write the result of the computation in Table XXV in the form $\sigma = 1.6429 \pm 0.0070$.

In the case of $\sigma(\sqrt{\beta_1})$ and $\sigma(\beta_2)$ no general formulæ are

given in Table XXX. The formulæ for the Normal Law, however, are usually a fair approximation, and are certainly so in the present case where we are dealing with data that is known to be distributed according to the Binomial Law, which does not differ in any very essential way from the Normal Law. If, then, we use the quantities $\sqrt{6/N}$ and $\sqrt{24/N}$ we get the results $\sqrt{\beta_1} = 0.1991 \pm 0.0151$ and $\beta_2 = 2.8778 \pm 0.0302$, respectively.

§ 111. *Control Charts*

This knowledge as to the nature of the distribution law obeyed by the various statistics that are commonly derived from statistical data has been used in a very elegant way in the construction of "control charts," the purpose of which is to reveal at a glance whether or not the universe from which a given sample was drawn was, or was not, of the type it was supposed to be.

For example, let us consider samples of 100 individuals each, taken from a supposedly normal universe with an average 0 and a standard deviation 2.6. We have said that each of the four quantities $\bar{n}$, σ, $\sqrt{\beta_1}$ and β_2 is distributed according to the Normal Law; and from Table XXX we find that their standard deviations, when gotten from samples of 100, are

$$\sigma(\bar{n}) = 0.260,$$
$$\sigma(\sigma) = 0.184,$$
$$\sigma(\sqrt{\beta_1}) = 0.245,$$
$$\sigma(\beta_2) = 0.490.$$

As the chance of a deviation exceeding the standard deviation by a factor of more than 2.5 is about 0.01, we may say that the chances are 100 to 1 that the experimental average will lie between ± 0.65, the standard deviation between 2.6 ± 0.46, the asymmetry between 0.0 ± 0.61, and the flatness between 3.00 ± 1.23, *if the experimental data really comes from the supposed universe.*

Suppose, now, that we were confronted with the problem of determining which of a large number of samples handed us really had come from such a universe, and which had not. Suppose that these samples were numbered 1, 2, 3, . . . , and that we computed from each of them the statistics mentioned.

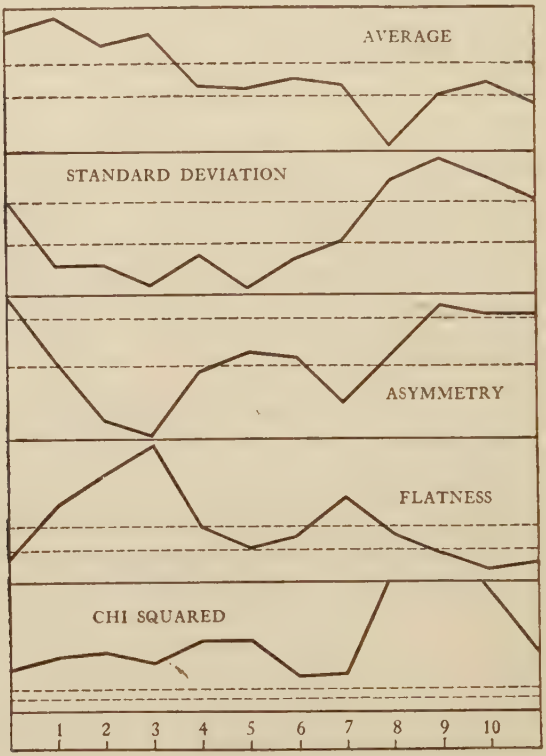


FIG. 38.—A TYPICAL CONTROL CHART.

Finally, suppose we plotted upon a chart, such as that shown in Fig. 38, the results obtained. The result would be, for each statistic, a jagged line drawn about the various expected values, somewhat as shown in the figure. We would obviously have no reason to suspect those samples for which the statistics were all very near their expected values; or indeed any which lay well within the band marked off with the dotted line, and which

is, in each case, the band beyond which a statistic has only a 1 in 100 probability of appearing. On the other hand, a sample for which certain statistics were outside this band would be decidedly questionable; and one for which all were outside, as is the case with number 3, would almost certainly not have arisen from a Normal Universe.

Such a chart as this is known as a Control Chart. It is used principally in factory inspection and kindred fields, where it is desired to know at a glance whether or not some extraneous influence is causing exceptional deviations from standard. Only a limited amount of computation is required to make the type of check to which we have referred. If desired, however, additional statistics could be included. For example, the use of χ^2 would give an even more sensitive check, and in this case the distribution function which it obeys is well known: it is merely the $P(> \chi^2)$ given in Appendix VIII.

PROBLEMS

1. A " direct advertising " sales campaign is under consideration, and a trial batch of 1000 circulars is sent out. It results in 19 favorable replies. Assuming that the Binomial Law sufficiently well represents the situation, state an upper and lower limit to the number of replies which may be expected from 100,000 circulars, it being understood that the limiting expectations are to be such that, beyond them, the chance of the trial batch showing 19 returns is less than 0.01.

2. How large a test batch must be used in Problem 1, in order that the upper and lower limits expected from a subsequent batch of 100,000 shall not differ by more than 100?

3. What is the least number of favorable replies in Problem 1 to assure a lower limit of expectation of 2 per cent?

4. Formulate a recommendation for a direct advertising campaign which requires 2 per cent to pay, covering the following points: 1000 circulars being standardized as the *first* test, (*a*) What shall be the " acceptance number " (that is, the number of favorable replies that shall be accepted as conclusive)?; (*b*) What shall be the " rejection number "?; (*c*) What procedure do you recommend in doubtful cases?

5. The results of 27 independent solutions of Problem 3, § 80 were as follows:

Number of Heads	Frequency of Occurrence (Individual Results)									
0	0	0	0	1	0	0	0	0	0	0
1	0	0	1	0	0	1	0	0	0	0
2	1	2	2	0	1	2	0	2	0	1
3	5	6	8	6	4	4	7	7	4	4
4	13	10	6	10	14	10	9	9	13	9
5	11	11	8	6	10	13	10	13	15	10
6	13	12	14	11	12	12	9	12	10	16
7	4	7	7	11	7	7	8	4	5	7
8	2	1	2	5	2	1	6	3	2	3
9	1	1	2	0	0	0	1	0	1	0
10	0	0	0	0	0	0	0	0	0	0
Average	5.10	5.10	5.20	5.42	5.14	5.02	5.48	5.00	5.18	5.38

Number of Heads	Frequency of Occurrence (Individual Results)									
0	0	0	0	0	0	0	0	0	0	0
1	2	1	0	1	1	0	0	1	0	0
2	1	3	2	3	4	2	3	1	3	3
3	7	12	3	11	3	8	9	7	6	4
4	16	7	9	7	8	8	7	7	7	7
5	10	6	20	10	11	12	8	13	14	15
6	7	11	7	8	11	11	12	9	12	11
7	4	6	7	5	8	8	5	8	4	6
8	2	4	2	5	4	1	5	3	1	4
9	1	0	0	0	0	0	1	1	3	0
10	0	0	0	0	0	0	0	0	0	0
Average	4.68	4.82	5.12	4.82	5.18	5.00	5.14	5.20	5.14	5.22

Number of Heads	Frequency of Occurrence (Individual Results)							Totals	Total for First 24	Total for Last 3
0	0	0	0	0	0	0	0	1	1	0
1	1	0	2	0	0	1	0	12	11	1
2	2	1	1	1	2	2	1	46	41	5
3	4	6	4	5	10	6	5	165	144	21
4	11	13	16	10	12	12	7	267	236	31
5	9	8	10	14	12	12	15	306	267	39
6	14	14	7	12	7	11	8	293	267	26
7	7	5	7	7	5	5	11	175	154	21
8	2	2	3	1	2	1	1	70	66	4
9	0	1	0	0	0	0	2	15	13	2
10	0	0	0	0	0	0	0	0	0	0
Average	5.10	5.12	4.90	5.12	4.70	4.80	5.42	5.092	5.107	4.873

Are the results contained in the column headed " Totals " consistent with the assumption that the pennies were unbiased?

6. Each individual column in this array may be regarded as an independent experiment to determine the average number of heads. The individual averages are all shown at the bottom of the columns. What is the probability of this array of averages, if the pennies were unbiased?

7. What is the probability that the total of the last three columns came from the same universe as the first 24?

8. What is the probability of the totals obtained, if the pennies were biased to the extent indicated by the average 5.092?

9. What would be the probability of obtaining the array of averages which we have presented, if the pennies were so biased?

10. Test the goodness of fit of the Poisson Law in the case of Table XV.

References for Outside Reading

I. *General:*

1. Rietz: *Mathematical Statistics.*
2. R. A. Fisher: *Statistical Methods for Research Workers.*
3. Arne Fisher: *Mathematical Theory of Probabilities.*

II. *Goodness of Fit:*

4. Pearson: *On the Criterion that a given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that it can be Reasonably Supposed to Have Arisen from Random Sampling,* Phil. Mag., Series V, Vol. 50 (1900), pp. 157–175.
5. R. A. Fisher: *On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P,* Journal of the Royal Statistical Society, Vol. 85 (1922), pp. 87–94; *The Conditions under which χ^2 Measures the Discrepancy between Observation and Hypothesis,* Journal of the Royal Statistical Society, Vol. 87 (1924), pp. 442–450.

III. *Sheppard's Corrections:*

6. Sheppard: *On the Calculation of the Most Probable Values of Frequency Constants, for Data Arranged According to Equidistant Divisions of a Scale,* Proceedings of the London Mathematical Society, Vol. 29 (1898), pp. 353–380.

IV. *Precision of Statistics:*

7. Pearson: *On the Probable Errors of Frequency Constants*, Biometrika, Vol. 2 (1903), pp. 273–281; *On the Probable Errors of Frequency Constants: Part II*, Biometrika, Vol. 9 (1913), p. 1; *On the Distribution of the Standard Deviation of Small Samples*, Biometrika, Vol. 10 (1915), p. 522.
8. "Student": *The Probable Error of a Mean*, Biometrika, Vol. 6 (1908), pp. 1–25.
9. R. A. Fisher: *Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population*, Biometrika, Vol. 10 (1915), pp. 507–521.

V. *Control Charts:*

10. Shewhart: *The Application of Statistics as an Aid in Maintaining Quality of a Manufactured Product*, Journal of American Statistical Association, Vol. 20 (1925), pp. 546–548; *Quality Control Charts*, Bell System Technical Journal, Vol. 5 (1926), pp. 593–603.

CHAPTER X

The Theory of Probability as Applied to Problems of Congestion

§ 112. *Introductory*

" Problems of Congestion " arise in any phase of industrial life in which demands for service arise from a multiplicity of sources acting more or less independently of one another. For example, the turnstiles through which passengers pass into a subway platform are used by numerous individuals who act independently of one another to a large degree, though perhaps influenced by common working hours to travel for the most part at certain peak hours. The demands made upon a cash-carrier system in a large store are " independent " in the same broad sense, though obviously they are influenced by peak shopping hours.

We have already had an example of one closely related problem. In § 86 we investigated the stock of dog-biscuit which should be carried by a grocery store under certain specified conditions. This, which we may for brevity refer to as the " warehouse problem," is in fact the simplest of all congestion problems. In it, the only question asked is: " How many demands will be made within a given time? " Obviously this question is also *part* of the problem raised by the turnstile or the cash-carrier, but not the whole problem. For the passenger who uses the turnstile does not remove it permanently from service, as the purchaser of dog-biscuit removes it permanently from stock. Instead, the turnstile is " returned to service " after the passenger is through with it: that is, *after a certain period called the holding-time.* This holding-time is a new element in the situation.

The fundamental turnstile problem therefore formulates itself as follows: " Knowing the expected number of demands per unit time and the expected holding-time, how many paths (turnstiles, say) must be provided in order that the proportion of persons inconvenienced shall not exceed a preassigned amount?"

This question is, however, still quite indefinite, for " inconvenienced " may mean a number of different things under different circumstances. In the case of the turnstile or cash-carrier it almost certainly means " delayed "; for the user does not disappear if no apparatus is available. If, however, we spoke of the number of chairs in a barber shop, a period of congestion would probably result in a loss of trade — and would therefore be to a certain degree its own cure, for the periods of congestion would obviously be shorter than if all the customers waited until served. This is probably no comfort to the barber, but we — and he — must face the facts nevertheless.

I doubt if my readers are likely to become barber-shop engineers; but there are other places where problems of an exactly analogous kind are faced in engineering experience. In telephone engineering, for instance, a certain number of trunk lines are provided between two exchanges, and when they are all busy the subscriber is given a " busy signal " which causes him to hang up his receiver and repeat his call later on. This is not quite a case of " lost traffic," for he very probably does repeat; but unless he repeats very soon — before the congestion is quite thoroughly cleared out — the length of such periods will be much the same as if he were to go away entirely.

We have, therefore, two quite fundamental divisions in this problem of congestion: a " delayed traffic " division and a " lost traffic " division. It is our purpose in the present chapter to indicate how the Theory of Probability can be applied to these two problems, and to two others which we shall explain as they arise.

As the methods of solution are the same no matter what the particular engineering application may be, it makes little

difference in what language we phrase our study. As telephone practice offers examples of widely diverse conditions, we shall choose it; and to make the study more understandable we give the following general explanation of the terms which we are to use:[1]

When a subscriber makes a call, it is designed for some particular person, and must therefore be steered toward that person, either by human intervention — as in manual practice — or by mechanical intervention — as in machine-switching (automatic) systems. Whatever performs this steering function we call a " switch."

Obviously a switch must pass a call on to something else. We call that something a " channel." Such a channel will ordinarily be one of a " group " performing identical functions — that is, any one of the group could accommodate our call. There may be other channels to which our call might have been assigned if it were going somewhere else — to a different office, say; these however are not part of our " group." There may also be other channels leading to the place we wish our call sent, and which are available to *other* subscribers though not to *us*. These also are not part of our " group." When we speak of a " group of channels " we shall mean a group each member of which is capable of performing identical functions for identically the same calls as any other.

Usually these calls can come from a number of sources. For example, if the group serves calling subscribers directly, there will usually be a number of subscribers capable of using any channel of the group in exactly the same way.[2] This is the " group of sources " corresponding to the " group of channels." In such cases, there must be something to locate a suitable channel and associate the calling source with it. Whatever performs this function, whether human or mechanical, we shall term a " switch."

[1] It need hardly be said that it is not intended as a description of a telephone exchange.

[2] But the source need not be a subscriber. It may be a switch to which he has already entrusted his call.

We need no further knowledge of telephony for the purposes of our discussion.

§ 113. *Notation*

The principal symbols used are the following:

n — the calling rate, that is, the expected number of calls *per source per hour.*

T — the expected duration of a call, measured in hours.

λ — the number of sources in our group.

ν — the number of channels in our group.

p — the probability that a given source (sometimes a given channel) is busy at a random instant of observation.

$P(j)$ — the probability that if a particular group is examined at a random instant it will be found to contain exactly j busy members.

Π — the probability of a call being lost by reason of insufficient equipment.

ϵ — the expected traffic density of our group; that is, the expected number of busy sources (or channels).

To any of these symbols will be affixed such subscripts and superscripts as are necessary to characterize the particular conditions to which they are applied.

§ 114. *General Assumptions*

We make the following assumptions once for all as to the nature of our problem:

Assumption 1 — The system is in statistical equilibrium; in other words, the probability of finding it in any specified condition is independent of the time at which it is examined.

It is quite true that no telephone exchange ever actually reaches a condition of statistical equilibrium. Its traffic varies from a light load at night to a peak sometime during the day, and then falls off again. During the time when the traffic is increasing, the probability of a large number of busy switches is also increasing and therefore varying with the time. On the other hand, when the traffic is decreasing, the probability of lost calls is also decreasing. It is not hard to see that the probability of losing calls always lags somewhat behind the traffic, reaching its peak shortly after the

peak load is reached, and its minimum shortly after the minimum load occurs. This was illustrated in a simple way by the results of Problems 8 and 9 of § 89. But when the periodic fluctuation of the traffic is sufficiently slow, the peak value of the probability is substantially the same as the probability of loss figured on the basis of statistical equilibrium with the traffic density at its peak value; and in such cases it is safe to make the assumption stated above, designing the exchange entirely for the conditions of busy hour traffic.

Assumption 2 — Connection of sources to channels and their disconnection therefrom is effected instantaneously.

This assumption is entirely tenable as long as the time consumed in the operations of connecting and disconnecting is small compared to the duration of the average conversation. In other cases an independent investigation is necessary; but to take into account such minor complications would only serve to obscure the main purpose of the present discussion.

Assumption 3 — The expected traffic density is the same for every source.

This assumption is justified only by the fact that it is difficult to make any other which more nearly agrees with practical conditions. It does not mean that each subscriber originates the same number of calls. It means that the number of seconds in the busy hour during which a source is expected to be busy is the same for all sources.

The assumption is evidently not satisfied in practice and some notion of the nature of the errors to which it leads is desirable. In an article on *The Theory of Telephone Probabilities Applied to Trunking Problems* in the Bell System Technical Journal for November, 1922, Mr. E. C. Molina has shown that it is on the side of safety, at least when the probability of congestion is small, as it usually is under operating conditions.

Assumption 4 — Busy sources make no calls.

If this assumption is ignored, formulæ will be obtained which give the same probability of loss for the same amount of traffic, regardless of whether the number of sources is less than or greater than the number of available channels. As an extreme example, consider the case of 20 sources, each originating five calls per hour, and the alternative case of five sources, each originating 20 calls per hour, the average duration of the calls being two minutes in each

case. If each of these groups is assigned ten channels, such a formula would say that the proportion of lost calls is the same in both cases. However, it is obvious from a common-sense standpoint that if there are only five sources they cannot make use of more than five channels. Hence in the second illustration no calls can possibly be lost, while it is possible for the twenty sources of the first illustration to want more than ten channels at once.

Assumption 5 — Either every channel which can serve a source S_1 can also serve S_2, or else no channel can serve them both.

We have really inferred this in the explanation given in § 112. There are cases in telephone practice (the "graded multiple" is a good example) which violate it.

Assumption 6 — The number of busy channels in a group is equal to the number of busy members in its group of sources, except that in case lost calls are not instantly wiped out, the number of busy sources may exceed the total number of channels. When this latter situation arises, all channels are busy.

This assumption is violated whenever the group of channels goes in only one of a number of different directions to which the sources have access; for obviously a source may be sending a call in one of these other directions. It is less often false in other engineering fields than in telephony.

§ 115. *Some Problems of Lost Traffic*

In order to illustrate the general principles involved we shall develop six formulæ for the probability of a call being lost. They illustrate well the extent to which shades of meaning must be carefully considered in dealing with such problems, arising as they do, on the one hand, from three slightly different assumptions as to how the traffic originates, and on the other hand from two as to what happens when a call is lost. The three which deal with the origination of calls are:

Assumption 7 — The probability that a particular source will originate a call during a given time interval is the same for every interval at the beginning of which it is idle. It is not in any way

influenced by the condition of its group of channels. (Alternative to Assumptions 8 and 9.)

Assumption 8 — The calls which are assigned to the group of channels are distributed individually and collectively at random.[1] *That is, the chance of the group being assigned a call during a test interval is independent of the state of either group.* (Alternative to Assumptions 7 and 9.)

The difference between these assumptions may be illustrated as follows: If a group of ten channels is accessible to fifteen subscribers and to them only, it seems scarcely reasonable to assert that the chance of a call being originated within a second is the same if all trunks are busy at the beginning of the second as it is if all trunks are idle; for when all trunks are idle there are three times as many possible sources as when all are busy. To make such an assertion implies that individual subscribers are more likely to call when the group is busy than when it is idle: and this in turn implies foreknowledge on their part. However, it is this assertion that is contained in Assumption 8. Unreasonable as it appears from this extreme illustration it will be found that it is often very near the fact when the sources of calls are not the subscribers themselves, but interoffice trunks and the like.[2]

A more reasonable assertion in case the subscriber is the source would be that the chance of a call originating during a short test interval is proportional to the number of idle subscribers; that is, in the case of the above illustration, that it is three times as great when all channels are idle as when all are busy. This is the condition implied by Assumption 7. It also is sometimes very near the truth, even when the sources are not subscribers' lines.

Assumption 9 — The probability of a call being assigned to the group of channels by some one of its sources is independent of the condition of either group, unless all sources are busy, in which case it is zero. In other words: calls are distributed individually and collectively at random, except that none is made when all channels are busy. (Alternative to Assumption 7 and 8.)

[1] In this connection, see the footnote accompanying the definition of "collectively at random" in § 84.

[2] It should also be noted that Assumption 8 implies, either that the chance of all sources being simultaneously busy is zero, or that Assumption 4 is violated. Any procedure which violates the latter assumption will be carefully avoided.

The discrepancies between the results given by 7 and 8 are frequently large enough for practical traffic densities that the use of the wrong formula would result either in inadequate or in extravagant installation. The results of 8 and 9 are generally so nearly alike as to be interchangeable in practice. The principal difference is, that while the use of 8 may, in extreme cases, require more channels than sources, 9 does not fall into this difficulty. Such a result is so absurd, however, that no one would put faith in it; so that the advantage which 9 appears to have in this respect is of doubtful value.

The assumptions dealing with what happens to lost traffic are:

Assumption 10 — If a call is lost because no channel is available, the source which made it nevertheless continues to demand service. If during this time a channel becomes available it will be seized and rendered unavailable for others for the entire period that would have been required for the call if it had been successful, though the call will still be regarded as lost. (Alternative to Assumption 11.)

Assumption 11 — If a call is lost by virtue of insufficient equipment its holding time is zero. (Alternative to Assumption 10.)

Assumption 10, of course, does not correspond to what actually takes place in telephone practice. If a call is unsuccessful, especially if the subscriber is informed of this fact, he is more likely to hang up his receiver quickly than otherwise. But there are problems to which Assumption 10 appears rigorously applicable. I believe certain types of fire-alarm apparatus are so designed that the sending mechanism continues to attempt to send an alarm for a fixed time, whether or not the alarm circuit is already in use.

In order to avoid a considerable amount of circumlocution these two conditions will be spoken of briefly as "lost calls held" and "lost calls cleared," and each will be considered under a separate topical heading.[1]

[1] "Lost calls held" must not be confused with "delayed calls"; for the latter stand by until served *and then consume a time equal to their holding time* — as would be the case if the fire-alarm mechanism were restrained from starting until its circuit was free.

§ 116. *The Elementary Probabilities; Lost Calls Held*

The two events of prime importance in a telephone system are the inception of a call and its termination. If the probability of the occurrence of each of these events is known under all circumstances, it should be possible to determine exactly how much equipment is needed. Hence the attack will be begun by evaluating them in accordance with the assumptions given above, using, to begin with, 7 and 10 as the particular pair of alternatives.

Suppose a source is tested at a certain instant and observed for a short time, dt, thereafter. At the moment when it is first tested it must be either idle or busy. The probability that it is busy has already been denoted by p; the probability that it is idle must therefore be $1 - p$. If the source is busy at the beginning of this interval, the only way in which it can originate a call is for the subscriber to close the call in progress and start another one during the interval. The chance of this happening can be made negligibly small by choosing a sufficiently short time interval dt. Then it will be true that if the source is busy at the beginning of dt, it cannot possibly originate a call before dt ends.

The chance that a source, idle at the beginning of dt, becomes busy before its end is denoted by $p_i(b)$, and the chance of the source originating a call during the time interval dt, assuming that nothing is known about its condition at the beginning of that interval by $p(b)$. Then by the rule for alternative compound probabilities the relationship

$$p(b) = (1 - p) \cdot p_i(b) + p \cdot p_b(b) \qquad (145)$$

may be written down at once. In words this relation expresses the logical proposition, that the chance of a source becoming busy during a random time interval must be equal to the chance that it is idle and being idle becomes busy, plus the chance that it is busy and being busy becomes busy again. The latter of these has already been said to be zero to the first order of the small quantity dt, and in accordance with the results of § 85 the random chance of a source originating a

call is obviously $p(b) = n\,dt$, to the same degree of approximation. Hence, inserting this value in equation (145), the chance of an idle line becoming busy is found to be

$$p_i(b) = \frac{n\,dt}{1 - p}. \tag{146}$$

This is one of the elementary probabilities.

If the chance of a source becoming *idle* during dt is considered, the logical proposition

$$p(i) = (1 - p) \cdot p_i(i) + p \cdot p_b(i) \tag{147}$$

is obtained, which expresses the fact that if a source becomes idle, it must either be idle and become idle again (the chance of which is negligibly small) or else it must be busy and become idle. However, the source obviously becomes idle just as often as it becomes busy, and if nothing whatever is known about the condition of the line at the beginning of the time interval dt, the chance of it becoming idle during that interval must be the same as the chance of it becoming busy. In other words, $p(i) = p(b) = n\,dt$. Inserting this value in (147) it is found that

$$p_b(i) = \frac{n\,dt}{p}. \tag{148}$$

If, instead of one source only, the entire group of λ sources are examined and j of them are found busy at the beginning of the interval, the chance that some one of these j busy sources will become idle during dt is just j times as great as the chance for one individual source. Also, if Assumption 7 is adopted the chance that one of the $\lambda - j$ idle sources will become *busy*, is just $\lambda - j$ times as great as the chance for one individual source. These latter probabilities are therefore [1]

$$\frac{j\,n\,dt}{p} \tag{149}$$

[1] There are two statements of a negative sort which it is worth while making with respect to these elementary probabilities. In the first place, passing from (146) to (150) *does not imply that the calling rate n is the same for all sources.* It is quite true

and

$$\frac{(\lambda - j)\, n\, dt}{1 - p}, \tag{150}$$

respectively. In both these formulæ p represents the chance that an instant of observation, chosen at random, finds the line busy. It is obviously equal to the proportion of the hour during which the line may be expected to be busy; that is, to nT. (See §§ 55–57.)

§ 117. *Introduction of the Assumption of Statistical Equilibrium; Lost Calls Held*

It is now possible to introduce the condition stated in Assumption 1; that is, to assert that the *probability* of the system being in any specified condition is the same at the end of the time interval dt as at its beginning.

Consider first a time at which all sources are idle. The chance of this condition existing is very small if the group has anything like the total amount of traffic which it can safely handle. Nevertheless, the condition *might* occur and therefore has some finite probability. This may be denoted by ${}^{\lambda}P'(0)$, the prime signifying the condition of lost calls held and the λ and 0 referring to the total number of sources in the group and to the number which are busy. Since there are λ idle sources the chance of *some one* of them becoming busy during

---

that if the j sources which are busy happen to be those which call with the least frequency, the chance of a call originating during the time dt is greater than that given by (150); while if they happen to be those which call with the greatest frequency the reverse is the case. To mention special cases such as these, however, implies special knowledge regarding the particular sources which are busy, and this, of course, is not allowable. If the properly weighted average of these probabilities is formed it is found to be identical with (150) above.

In the second place *it is nowhere assumed that all calls are of the same length.* What *is* assumed is what is stated in Assumption 3 — that p is the same for every source. Although the method of derivation which has been used is entirely free from the implication of equal holding times, the assumption of equality has been so frequently employed by other writers, even when their results could just as well have been obtained without it, that it seems well to point out that it is not here involved.

a time dt is $\lambda n\, dt/(1-p)$, and the probability that *none* will become busy during this length of time is

$$1 - \frac{\lambda n\, dt}{1-p}.$$

Hence the chance that all the sources are idle both at the beginning and at the end of the interval is

$$\left(1 - \frac{\lambda n\, dt}{1-p}\right) {}^{\lambda}P'(0).$$

This is not, however, the total probability that none of the sources is busy at the end of this interval, for it *might* happen that at the beginning of the interval one source was busy, and that during the interval this source became idle. Denoting the probability that exactly one source is busy by ${}^{\lambda}P'(1)$, and noting that the probability of this source becoming idle is, by formula (149), dt/T, it is easily seen that the probability of it being busy at the beginning of the interval and idle at the end is

$$\frac{dt}{T} \cdot {}^{\lambda}P'(1).$$

There are other things which might conceivably happen during dt which would leave the entire group of sources idle when this interval closes. For example, two sources might be busy and both of them become idle. But if the chance of a particular source becoming idle is dt/T, the chance that both of two busy sources will become idle is the square of this, and is therefore of the second order in the very small quantity dt. In fact, a little consideration serves to show that the probability of any one change taking place in the condition of the set of sources is of the first order in dt; the probability of any two changes is of the second order; and so on. Since quantities of the second or higher orders in dt are so small as to be negligible, it follows that there is no need of considering the possibility of more than one such change taking place. Hence, to the first order of small quantities, the probability of all the sources of the system being idle at the end of the

interval dt is the sum of the probability that all were idle at its beginning and remained so, and of the probability that one only was busy at the start and that this one became idle. These quantities having already been found, the probability of all sources being idle at the end of dt is easily seen to be

$$\left(1 - \frac{\lambda n\, dt}{1 - p}\right) {}^{\lambda}P'(0) + \frac{dt}{T}\, {}^{\lambda}P'(1).$$

To accord with the assumption of statistical equilibrium this must be the same as ${}^{\lambda}P'(0)$. Forming the equation to which this fact leads and making a few obvious cancellations, it is found that

$$\frac{\lambda n}{1 - p}\, {}^{\lambda}P'(0) = \frac{1}{T}\, {}^{\lambda}P'(1). \tag{151}$$

A similar argument may be applied to the case where every source is busy. The probability that some one of the sources will become idle being $\lambda\, dt/T$, it follows that the probability of none of them becoming idle is $1 - \lambda\, dt/T$. If, therefore, the probability that λ sources are busy at the beginning of the interval is denoted by ${}^{\lambda}P'(\lambda)$, the chance that they will be busy both at the beginning and at the end of the interval is

$$\left(1 - \frac{\lambda\, dt}{T}\right) {}^{\lambda}P'(\lambda). \tag{152}$$

There is only one other way in which λ sources of the system may be busy at the end of the interval without more than one event taking place in the meantime. This occurs in case $\lambda - 1$ sources are busy at the start and the remaining one becomes busy before the interval ends. The probability of this is the product of the probability ${}^{\lambda}P'(\lambda - 1)$ that exactly $\lambda - 1$ sources are busy at the beginning of the interval, and the probability $n\, dt/(1 - p)$ that the remaining one becomes busy before it ends. Adding this product to (152), and remembering that the result must be ${}^{\lambda}P'(\lambda)$ the equation

$$\frac{n}{1 - p}\, {}^{\lambda}P'(\lambda - 1) = \frac{\lambda}{T}\, {}^{\lambda}P'(\lambda). \tag{153}$$

is obtained.

In general, the condition of j and only j sources busy at the end of the interval may occur in either one of three ways:

(a) By exactly j being busy at the beginning of the interval and no calls being originated or discontinued during it, the probability of which is

$$\left(1 - \frac{\lambda - j}{1 - p}\, n\, dt - \frac{j}{T}\, dt\right) {}^{\lambda}P'(j);$$

(b) By exactly $j - 1$ being busy at the beginning of the interval and one new call originating, the probability of which is

$$\frac{\lambda - j + 1}{1 - p}\, n\, dt\, {}^{\lambda}P'(j - 1);$$

(c) By exactly $j + 1$ being in progress at the beginning of the interval and one being completed, the probability of which is

$$\frac{j + 1}{T}\, dt\ {}^{\lambda}P'(j + 1).$$

By summing these three terms to get the complete probability of exactly j busy sources at the end of the interval, and setting this probability equal to ${}^{\lambda}P'(j)$, an equation is obtained which may easily be reduced to the form

$$\left(\frac{\lambda - j + 1}{1 - p}\, n\right) {}^{\lambda}P'(j - 1) - \left(\frac{\lambda - j}{1 - p}\, n + \frac{j}{T}\right) {}^{\lambda}P'(j) + \left(\frac{j + 1}{T}\right) {}^{\lambda}P'(j + 1) = 0. \qquad (154)$$

§ 118. *The Probability Formulæ Corresponding to Assumptions 7 and 10*

If the equations (151), (153), and (154) were linearly independent they would be sufficient to determine the value of the probabilities ${}^{\lambda}P'(j)$ for each j from 0 to λ; for there are exactly $\lambda + 1$ of these probabilities and there are exactly $\lambda + 1$ equations corresponding to them. It so happens, however, that they are not linearly independent and an additional

equation is necessary to solve the problem. This is readily obtained by remembering that

$$\sum_{j=0}^{\lambda} {}^{\lambda}P'(j) = 1. \tag{155}$$

Solving these equations, which may be done by the theory of determinants,[1] it is found that

$${}^{\lambda}P'(j) = C_j^{\lambda} p^j (1 - p)^{\lambda - j}. \tag{156}$$

This equation gives with absolute accuracy the probability of exactly j busy sources out of a total of λ at an arbitrary instant when a test is made, provided Assumptions 7 and 10 are satisfied. It is, therefore, the probability that exactly j subscribers will wish to use the group of channels simultaneously.

The formula itself is the usual Binomial Law for the probability of an event happening j times in λ independent trials, if the probability of success in a single trial is p. The problem could have been so phrased that the answer would have been apparent at once: the longer method was adopted because it emphasizes the underlying hypotheses, and leaves no doubt as to the exact meaning of the answer when obtained.

To find the probability of a call being lost we make use of the following argument: If we choose an interval at random, and observe the system during this interval, a call may or may not occur. If it does, it may or may not be lost; but since the interval has been chosen at random, without regard for the state of the system, " the probability that it is lost if it occurs " is just exactly the thing that we mean by the words " the probability of loss."

In the form in which we have stated it, however, this is a conditional probability, and to it the formula (20) may be applied at once if we let the symbol AB mean " a call occurs

[1] An alternative method of solution is: to find ${}^{\lambda}P'(1)$ in terms of ${}^{\lambda}P'(0)$ from (151); then by writing $j = 1$ in (154) to find ${}^{\lambda}P'(2)$ in terms of ${}^{\lambda}P'(0)$; next by writing $j = 2$ in (154) to find ${}^{\lambda}P'(3)$ in terms of ${}^{\lambda}P'(0)$; and so on. After every ${}^{\lambda}P'(j)$ has been so expressed, ${}^{\lambda}P'(0)$ may be found from (155).

and is lost," the symbol, A, " a call occurs," and the symbol B, " it is lost."

As for the chance of a call occurring and being lost — that is, $P(AB)$ — that is just

$$\sum_{j=\nu}^{\lambda} {}^{\lambda}P'(j)\,\frac{(\lambda - j)\,n\,dt}{1-p} = \lambda n\,dt \sum_{j=\nu}^{\lambda-1} C_j^{\lambda-1}\,p^j(1-p)^{\lambda-j-1} \qquad (157)$$

for if more than ν sources are busy during our interval dt and a call occurs it will of necessity be lost. And as for the chance of a call occurring, — that is, $P(A)$ — that is just $\lambda n\,dt$. So substituting these values in (20) and making certain simple rearrangements we get for $P(B)$, or ${}^{\lambda}\Pi'_{\nu}$, the form

$${}^{\lambda}\Pi'_{\nu} = \left(\frac{\lambda - \epsilon}{\lambda}\right)^{\lambda-1} \sum_{j=\nu}^{\lambda-1} C_j^{\lambda-1} \left(\frac{\epsilon}{\lambda - \epsilon}\right)^j \qquad (158)$$

where ϵ, or λp, is the expected traffic density in the group.

§ 119. *The Probability Formulæ Corresponding to Assumptions 8 and 10*

Formulæ (156) and (158) are more complicated than is necessary for many purposes, for it frequently happens that the traffic arises from a very large number of sources, each one of which is busy but a small fraction of the time. In such cases the number of idle sources — and therefore the chance of a new call — is substantially the same at every instant, and we should expect to be able to find a suitable formula in a simpler form. Indeed we find, by allowing the number of sources λ to increase indefinitely without changing either ϵ or ν, that (158) approaches the simpler expression

$${}^{\infty}\Pi'_{\nu} = e^{-\epsilon} \sum_{j=\nu}^{\infty} \frac{\epsilon^j}{j!}. \qquad (159)$$

This formula is much used in computing trunk groups.

The corresponding formula for the probability of exactly

j busy sources is obtained by taking the limit of (156), the result being, as we have seen in § 83,

$$^{\infty}P'(j) = \frac{\epsilon^j e^{-\epsilon}}{j!}. \qquad (160)$$

which is just the familiar Poisson Formula.

The manner in which (159) has been derived suggests that it is only accurate when the number of sources greatly exceeds the number of channels to which they have access. As a matter of fact, this is true if each source is independent of the rest as required by Assumption 7. However, the usefulness of (159) is actually much broader than this statement would imply, as can be shown by placing it upon a slightly different foundation, as follows:

Since $\epsilon = \lambda nT = \lambda p$, it follows that as λ is increased indefinitely, nT and p must each decrease according to the law $nT = p = \epsilon/\lambda$. Inserting this in (150) we find that the chance of a call being originated when j sources are busy is

$$\left(1 - \frac{j}{\lambda}\right) \frac{\epsilon}{1 - \frac{\epsilon}{\lambda}} \frac{dt}{T},$$

a quantity which approaches the limit $\epsilon\, dt/T$ as λ increases indefinitely. Since this limit is independent of j it follows that *formula (159) corresponds to Assumption 8, that is, to the case where the calls are distributed individually and collectively at random.*

From a practical standpoint Assumption 8 is inconsistent with a limited number of sources, for in practice it must always be true that the chance of a new call being originated when all the sources are busy is zero; and it is therefore dependent on the number of busy sources to just that extent.

This practical difficulty is reflected in the theory in the sense that the combination of Assumptions 8 and 10 is inconsistent with Assumption 4 unless the number of sources is infinite. For the purpose of this paragraph, therefore, Assumption 4 may be regarded as ignored. The same difficulty will not arise when Assumption 8 is

combined with Assumption 11 unless the number of channels is at least as great as the number of sources.

Practically speaking, these difficulties in harmonizing our assumptions are unimportant unless the chance of all sources being busy simultaneously is quite large. Moreover, it is actually true that (159) and (160) are extremely valuable in many cases where the number of sources exceeds the number of channels by a sufficiently wide margin.

§ 120. *The Probability Formulæ Corresponding to Assumptions 9 and 10*

The practical analogue of Assumption 8 in the case of a limited number of sources is Assumption 9, which states that so long as any sources are idle the chance of a new call being originated is independent of their number, but that as soon as all sources become busy the chance of a new call being originated drops to zero. If the method of computation which has been used in obtaining formula (158) is applied to this set of assumptions the results

$$P'(j) = \frac{\frac{\epsilon^j}{j!}}{\sum_{j=0}^{\lambda} \frac{\epsilon^j}{j!}} \tag{161}$$

and

$$\Pi'_\nu = \frac{\sum_{j=\nu}^{\lambda-1} \frac{\epsilon^j}{j!}}{\sum_{j=0}^{\lambda} \frac{\epsilon^j}{j!}} \tag{162}$$

are obtained.

It can be shown that in most instances these formulæ give approximately the same values as those obtained from (159) and (160). Practical conditions usually require that the probability of loss shall be small. This means, of course, that the terms in the numerator of (162) must be small, and it is easy to see that if this is true the difference between the denominator and a similar expression summed from zero to infinity is negligibly small. The latter expression, however, is the series expansion for e^{ϵ}. If this approximation is sub-

stituted for the denominator, (161) immediately becomes identical with (160).

Likewise if it is true that λ is much larger than ν the difference between the numerator of (162) and a similar expression summed from ν to infinity will be negligible and (162) will reduce to (159). In other words, (161) is always sensibly equal to (160) under practical conditions, and (162) is approximately equal to (159) except when the number of sources is very nearly the same as the number of channels. These qualitative assertions will be given quantitative illustration in §§ 126 and 127.

The one vital difference between (162) and (159) is that (159) does not depend upon λ at all and therefore gives a finite probability of loss even when the number of channels exceeds the number of sources — an absurd result to which (162) does not lead.

This absurdity is merely the practical manifestation of the remark made in § 119: that Assumption 8 is not strictly tenable in any case where the number of sources is limited.

The present section and the two which precede it contain formulæ corresponding to the conditions of lost calls held, both when the sources of calls are assumed to be independent and when they are assumed to be dependent upon one another in such a way that the chance of a call originating is not influenced by the number of busy sources. It is necessary next to obtain analogous results for the condition of lost calls cleared.

§ 121. *The Elementary Probabilities; Lost Calls Cleared*

A careful consideration of the derivation of (146) shows that the form of this equation is not affected by shifting to the assumption of lost calls cleared. The value of p, however, is somewhat altered, due to the fact that unsuccessful calls contribute nothing to the busy time of the sources. Hence instead of $p = nT$ we now have $p = (1 - \Pi)nT$.

In the development of the second elementary probability, the logical proposition stated in (147) is no longer true, since it can no longer be asserted that the probability of an idle source becoming busy and idle again during the interval dt is negligible; for if such an idle source were to originate a call at a time when there was no available channel to receive it, it would instantly become idle again. Thus, in effect, an idle

source becomes idle, and $p_i(i)$ is not zero. Instead, $p_i(i)$ is now equal to the probability that all channels are occupied during dt, and that our source, which is idle, originates a call. We find at once

$$p_i(i) = \Pi \frac{n\,dt}{1 - p}.$$

Inserting this in (147), we get

$$p_b(i) = \frac{(1 - \Pi)\,n\,dt}{p};$$

and then remembering that p is now $(1 - \Pi)\,nT$, we are again led to the same formula dt/T as before.

It is obvious from a common-sense standpoint that the progress of a successful call, after its connection is established, should be in no way influenced by unsuccessful calls. In particular, the probability of termination and the holding time should be unaltered, whatever becomes of the unsuccessful calls. It would seem apparent, therefore, that if the chance of a busy line becoming idle is expressed in terms of dt and T only, the formula which results must be valid either for lost calls held or for lost calls cleared. This would establish the validity of (148) even if the method by which it was originally derived had introduced Assumption 9.

§ 122. *The Probability Formulæ Corresponding to Assumptions 7 and 11*

Having seen that both elementary probabilities are expressible in the same form as in the preceding case, it may be inferred at once that the form of equations (151) and (154) remains unchanged. This inference is borne out by an independent investigation. There is this difference in the circumstances, however: that, whereas in the preceding case the maximum number of sources which might be simultaneously busy was λ, in the present instance the number is ν. Equations (153) and (155) must therefore be reconsidered.

Suppose the probability of exactly ν channels being simultaneously busy is ${}^{\lambda}P''(\nu)$, the λ and ν having the same significance as before and the double prime relating to the condition

of lost calls cleared. If the system is to be in this condition at the end of a short interval dt, it may either have been so at the beginning of the interval and remained unchanged, or else it may have had just one idle channel at the beginning of the interval, this one becoming busy meanwhile. Taking both of these possibilities into account and introducing the principle of statistical equilibrium in exactly the same fashion as has been done above, it may be easily seen that (153) must be replaced by

$$\frac{\nu}{T}\,{}^{\lambda}P''(\nu) = \frac{\lambda - \nu + 1}{1 - p}\,n\,{}^{\lambda}P''(\nu - 1).$$

Similarly the sum of the probabilities of each number of busy sources from 0 to ν must be equal to unity, since it is impossible for this number to exceed λ. This gives

$$\sum_{j=0}^{\nu} {}^{\lambda}P''(j) = 1,$$

which takes the place of (155).

Having thus obtained the necessary independent equations, their solution can be carried out very easily by the use of determinants. The result is

$${}^{\lambda}P''(j) = \frac{C_j^{\lambda}\left(\frac{nT}{1-p}\right)^j}{\sum_{j=0}^{\nu} C_j^{\lambda}\left(\frac{nT}{1-p}\right)^j}. \tag{163}$$

It is desirable to express this formula in terms of the traffic density of the group. This traffic density is, as before, $\epsilon = \lambda nT$; whence

$$p = \frac{(1 - \Pi)\,\epsilon}{\lambda}.$$

Hence in terms of ϵ, (163) becomes

$${}^{\lambda}P''(j) = \frac{C_j^{\lambda}\left(\frac{\epsilon}{\lambda - (1-\Pi)\,\epsilon}\right)^j}{\sum_{j=0}^{\nu} C_j^{\lambda}\left(\frac{\epsilon}{\lambda - (1-\Pi)\,\epsilon}\right)^j}. \tag{164}$$

The probability of loss is obtained by the same argument as in § 118. The chance of a call being originated during a short interval of observation is $\lambda\, n\, dt$ as before. The chance of a call being lost, however, is much simpler, since it is now impossible for more than ν sources to be busy simultaneously. It is

$$ {}^{\lambda}P''(\nu)\,\frac{\lambda-\nu}{1-p}\,n\,dt. $$

The ratio of these two quantities is the probability of loss It is easily seen to be

$$ {}^{\lambda}\Pi''_{\nu} = \left(\frac{\lambda-\nu}{\lambda}\right)\frac{{}^{\lambda}P''(\nu)}{1-p} = \frac{1}{1-p}\,\frac{C_{\nu}^{\lambda-1}\left(\dfrac{nT}{1-p}\right)^{\nu}}{\sum\limits_{j=0}^{\nu} C_{j}^{\lambda}\left(\dfrac{nT}{1-p}\right)^{j}}. $$

A less obvious form, but one which is more convenient for computation is

$$ {}^{\lambda}\Pi''_{\nu} = \frac{C_{\nu}^{\lambda-1}\left(\dfrac{\epsilon}{\lambda-(1-\Pi)\,\epsilon}\right)^{\nu}}{\sum\limits_{j=0}^{\nu} C_{j}^{\lambda-1}\left(\dfrac{\epsilon}{\lambda-(1-\Pi)\,\epsilon}\right)^{j}}. \qquad (165) $$

§ 123. *The Probability Formulæ Corresponding to Assumptions 8 and 11*

Formulæ (164) and (165) are analogous to (156) and (158). From them others may be derived which are appropriate when the sources are independent and their number greatly exceeds the number of channels, or when the sources, though not very numerous, are so related that as more and more of them become busy the individual calling rates of those which remain idle increase at a rate which just neutralizes their decrease in number. This is done by taking the limits of (164) and (165) as λ becomes infinite, just as was done in § 119. In this way

formulæ analogous to (159) and (160) are obtained. They are

$$^{\infty}P''(j) = \frac{\frac{\epsilon^j}{j!}}{\sum_{j=0}^{\nu} \frac{\epsilon^j}{j!}} \tag{166}$$

and

$$^{\infty}\Pi''_{\nu} = \frac{\frac{\epsilon^{\nu}}{\nu!}}{\sum_{j=0}^{\nu} \frac{\epsilon^j}{j!}}. \tag{167}$$

As has been said in connection with formulæ (159) and (160), Assumption 8 is not tenable if it is possible for all sources to be busy simultaneously, for in this case there are no idle sources left to originate calls. This manifests itself in the fact that (167), like (159), gives a finite probability of loss, even when $\lambda < \nu$. For this reason, Assumption 4 must be ignored in developing (167), if $\lambda \leqq \nu$.

§ 124. *The Probability Formulæ Corresponding to Assumptions 9 and 11*

No change is made in either (166) or (167) when Assumption 8 is replaced by Assumption 9, unless $\lambda \leqq \nu$. However, if $\lambda < \nu$ they become, respectively,

$$P''(j) = \frac{\frac{\epsilon^j}{j!}}{\sum_{j=0}^{\lambda} \frac{\epsilon^j}{j!}}; \tag{168}$$

and

$$\Pi''_{\nu} = 0. \tag{169}$$

These formulæ are identical with (161) and (162).[1] That this

[1] When $\lambda \leqq \nu$ the upper index of summation in the numerator of (162) is less than the lower index, so that the formula is meaningless as written in § 120. Its true value, however, is zero, as listed in Table XXXII.

TABLE XXXI

SCHEMATIC REPRESENTATION OF NOTATION

Assumptions as to Origination of Calls	Assumptions as to Treatment of Lost Calls	
	Lost Calls Cleared (Assumption 11)	*Lost Calls Held* (Assumption 10)
Sources Independent (Assumption 7)	I: ${}^{\lambda}P''(j)$ ${}^{\lambda}\Pi''_{\nu}$	II: ${}^{\lambda}P'(j)$ ${}^{\lambda}\Pi'_{\nu}$
Calls occur individually and collectively at random (Assumption 8)	III: ${}^{\infty}P''(j)$ ${}^{\infty}\Pi''_{\nu}$	IV: ${}^{\infty}P'(j)$ ${}^{\infty}\Pi'_{\nu}$
Calls occur individually and collectively at random, unless all sources are busy; then none occur (Assumption 9)	V: $P''(j)$ Π''_{ν}	VI: $P'(j)$ Π'_{ν}

Note. — The Roman numerals are for the purpose of identification in connection with the curves which follow.

Formulæ I, II, III and IV are known, respectively, by the names Engset, Binomial, Erlang and Poisson.

is to be expected is obvious, since when no calls are lost, what happens to lost calls is immaterial.

§ 125. *Recapitulation of Formulæ*

Formulæ have now been obtained corresponding to each of the six alternative pairs of assumptions, and it is desirable to collect them together for purposes of reference. This is done in Tables XXXI to XXXIII.

Table XXXI represents schematically the relationship of the various formulæ to the sets of assumptions upon which they are based; while Tables XXXII and XXXIII give the formulæ themselves.

In developing these probabilities in the preceding pages there has been no need to write down the circumstances under

TABLE XXXII

PROBABILITY OF LOSS

Assumptions [1]	Formulæ		Reference Number
7 and 11	${}^{\lambda}\Pi''_{\nu} = 0$	$\lambda \leqq \nu$	(165)
	${}^{\lambda}\Pi''_{\nu} = \dfrac{C_{\nu}^{\lambda-1}\left(\dfrac{\epsilon}{\lambda-(1-\Pi)\epsilon}\right)^{\nu}}{\sum_{j=0}^{\nu} C_{j}^{\lambda-1}\left(\dfrac{\epsilon}{\lambda-(1-\Pi)\epsilon}\right)^{j}}$	$\lambda > \nu$	
7 and 10	${}^{\lambda}\Pi'_{\nu} = 0$	$\lambda \leqq \nu$	(158)
	${}^{\lambda}\Pi'_{\nu} = \left(\dfrac{\lambda-\epsilon}{\lambda}\right)^{\lambda-1} \sum_{j=\nu}^{\lambda-1} C_{j}^{\lambda-1}\left(\dfrac{\epsilon}{\lambda-\epsilon}\right)^{j}$	$\lambda > \nu$	
8 and 11 (4 ignored if $\lambda \leqq \nu$)	${}^{\infty}\Pi''_{\nu} = \dfrac{\dfrac{\epsilon^{\nu}}{\nu!}}{\sum_{j=0}^{\nu} \dfrac{\epsilon^{j}}{j!}}$	$\lambda \geqq 0$	(167)
8 and 10 (4 ignored if λ is finite)	${}^{\infty}\Pi''_{\nu} = e^{-\epsilon} \sum_{j=\nu}^{\infty} \dfrac{\epsilon^{j}}{j!}$	$\lambda \geqq 0$	(159)
9 and 11	$\Pi''_{\nu} = 0$	$\lambda \leqq \nu$	(169)
	$\Pi''_{\nu} = {}^{\infty}\Pi''_{\nu}$	$\lambda \wedge \nu$	
9 and 10	$\Pi'_{\nu} = 0$	$\lambda \leqq \nu$	(162)
	$\Pi'_{\nu} = \dfrac{\sum_{j=\nu}^{\lambda-1} \dfrac{\epsilon^{j}}{j!}}{\sum_{j=0}^{\lambda} \dfrac{\epsilon^{j}}{j!}}$	$\lambda > \nu$	

[1] Unless explicitly stated all assumptions from 1 to 6 are used.

which they take the value zero. This was evident from the context. In the tables, however, these limiting conditions are given in order that no ambiguity may be involved.

§ 126. *Numerical Comparison of Formulæ; The Dependence of the Probability of Loss upon the Number of Sources when the Traffic Density of the Group is Held Constant*

In order to gain some conception of the magnitude of the differences between these various formulæ, it is desirable to present a few numerical examples which illustrate the essential points of their behavior.

In the first place, the extent to which they depend upon

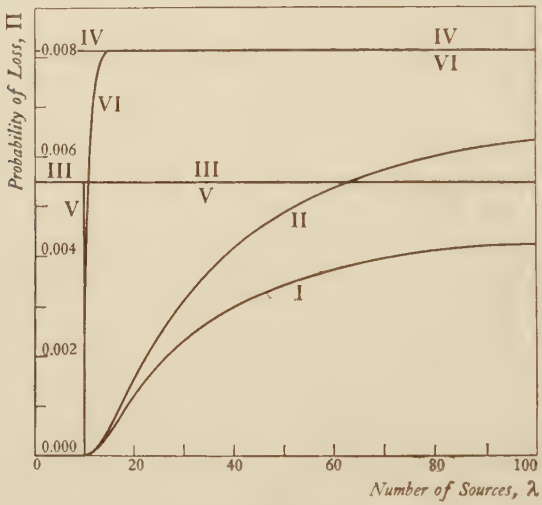


Fig. 39.—Comparison of Various Formulæ for Probability of Loss when the Traffic Density is Held Constant.

the number of sources may be considered. To illustrate this point a group of ten channels [1] is chosen and it is assumed that this group receives its traffic, sometimes from a few busy sources, sometimes from many relatively idle ones, but always

[1] Attention should be called to the fact that in this and the following sections the particular numerical values chosen are such as to give quite appreciable differences between the various formulæ. It would be a mistake to infer that the differences are always of this order of magnitude. As a matter of fact they may be either larger or smaller. It may be stated as a rough general rule — though this rule like most others has its exceptions — that where the groups of channels are large, the differences will be smaller than those here obtained, and vice versa.

TABLE XXXIII

PROBABILITY OF CONGESTION j

Assumptions [1]	Formulæ		Reference Number
7 and 11	$${}^{\lambda}P''(j) = \frac{C_j^{\lambda}\left(\frac{\epsilon}{\lambda-(1-\Pi)\epsilon}\right)^j}{\sum_{j=0}^{\nu} C_j^{\lambda}\left(\frac{\epsilon}{\lambda-(1-\Pi)\epsilon}\right)^j}$$	$j \leqq \nu$	(164)
	${}^{\lambda}P''(j) = 0$	$j > \nu$	
7 and 10	$${}^{\lambda}P'(j) = \left(\frac{\lambda-\epsilon}{\lambda}\right)^{\lambda} C_j^{\lambda}\left(\frac{\epsilon}{\lambda-\epsilon}\right)^j$$	$j \geqq 0$	(156)
8 and 11 (4 ignored if $\lambda < \nu$)	$${}^{\infty}P''(j) = \frac{\frac{\epsilon^j}{j!}}{\sum_{j=0}^{\nu}\frac{\epsilon^j}{j!}}$$	$j \leqq \nu$	(166)
	${}^{\infty}P''(j) = 0$	$j > \nu$	
8 and 10 (4 ignored if λ is finite)	$${}^{\infty}P'(j) = e^{-\epsilon}\frac{\epsilon^j}{j!}$$	$j \geqq 0$	(160)
9 and 11	$P''(j) = P'(j)$	$\lambda \leqq \nu$	(168)
	$P''(j) = {}^{\infty}P''(j)$	$\lambda \geqq \nu$	
9 and 10	$$P'(j) = \frac{\frac{\epsilon^j}{j!}}{\sum_{j=0}^{\lambda}\frac{\epsilon^j}{j!}}$$	$j \leqq \lambda$	(161)
	$P'(j) = 0$	$j > \lambda$	

[1] Unless explicitly stated all assumptions from 1 to 6 are used.

in such a way that the traffic density is 4. The results plotted against λ, the number of sources, are shown in Fig. 39, each curve corresponding to one of the formulæ of Table XXXII.

The two formulæ corresponding to the assumption of inde-

pendent sources (the top row in the scheme of Table XXXI) give the Curves I and II. These coincide with the λ-axis so long as $\lambda \leqq \nu$, and rise gradually as λ increases beyond this value.

The formulæ corresponding to Assumption 8 give horizontal lines, as is to be expected from the fact that the assumption implies independence of λ. They are designated III and IV. That they are asymptotic to the Curves I and II is evident from the fact that (167) and (159) were obtained as limiting cases of (165) and (158). They do not approach the λ-axis even when $\lambda < \nu$, a fact which is merely the graphical equivalent of the statement already made — that they give a finite probability of loss even when the number of channels exceeds the number of sources.

Curves V and VI, which correspond to the third row in Table XXXI, occupy an intermediate position. They coincide with the λ-axis for $\lambda < \nu$ and thus avoid the absurd results to which formulæ (167) and (159) give rise. Indeed, the purpose of the modification of Assumption 8 contained in Assumption 9 was exactly to avoid this absurdity.

For all values of λ which exceed ν, Curve V coincides with Curve III. That this is as it should be is seen from the fact that when λ exceeds ν it is not possible for all the sources to be simultaneously busy, and hence the modification of Assumption 8 plays no part whatever. Curve VI, on the other hand, while it rises more steeply than II, does not jump abruptly from zero to its maximum value but approaches the latter asymptotically. The reason for this lies in the fact that if lost calls are held, there can be more than ν in progress at one time. In this case, more than one must clear before a successful call can be made. Thus, those calls which fail still produce a " hang-over " effect, which interferes with the chance of success of other calls. This " hang-over " becomes greater and greater as the number of sources is increased. Indeed, it is this effect which is responsible for the fact that all of the curves corresponding to the first column of Table XXXI show greater probabilities of loss than do the analogous curves of the second column.

It may seem surprising at first thought that the "hang-over" effect should ever produce an increase in the number of lost calls as great as that which is necessary to account for the difference between Curves III and IV. In fact, Curve III says that, if lost calls are cleared, only about one-half of one per cent of the calls are lost, and Curve IV says that this small proportion, if held instead of cleared, is capable of increasing the proportion of loss by about 50 per cent. It should be remembered in this connection, however, that the very fact that calls are lost implies that they are originated at a time when the system is already congested. Therefore, unless they are quickly disposed of, a very few of them may lengthen the period of congestion to a considerable extent and increase the proportion of loss correspondingly.

That there is no "hang-over" effect when λ exceeds ν by 1 is evident from a common-sense point of view. Hence the modified Assumption 8 should give exactly the same results regardless of whether lost calls are held or cleared. In other words, Curves III, V and VI should all cross at the value $\lambda = 11$. That they do so is evident from the figure, as well as from the fact that in this case (167), (169) and (162) are all identical.

§ 127. *Numerical Comparison of Formulæ; The Dependence of the Allowable Traffic Density upon the Number of Sources, when the Proportion of Loss is Fixed*

The curves of Fig. 39 show very satisfactorily the essential differences between the results to which our various combinations of assumptions lead. They are open to the objection, however, that they give an exaggerated idea of the practical importance of these differences. Ordinarily probability formulæ are not used, as is done here, to compute the proportion of loss which corresponds to a preassigned amount of traffic, but for the converse purpose of computing the maximum allowable traffic density when the allowable proportion of loss is known. Since small changes in the traffic density produce large changes in the proportion of loss, formulæ which give widely different results when used in the former way may agree surprisingly well when used in the latter.

In order that no such erroneous impressions may be produced, curves showing the traffic density corresponding to a

loss of one call in one hundred are shown in Fig. 40. As before, a group of ten channels is considered and the number of sources is varied through the range from 0 to 100. The different curves correspond to the six alternative formulæ of Table XXXII.

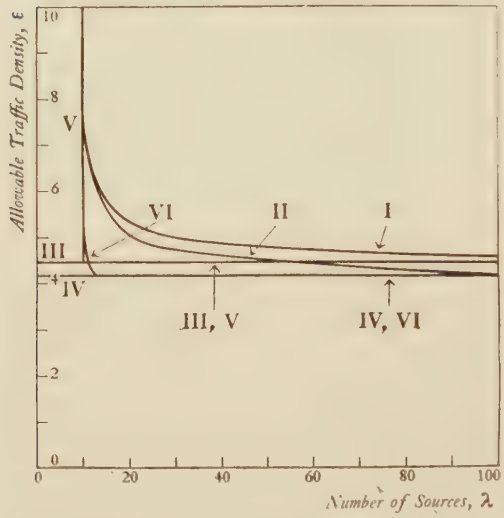


FIG. 40.—COMPARISON OF VARIOUS FORMULÆ FOR THE ALLOWABLE TRAFFIC DENSITY WHEN THE PROBABILITY OF LOSS IS HELD CONSTANT.

Formulæ (167) and (159) again lead to the straight lines III and IV, which extend unbroken even when $\lambda < 10$. Formula (169) leads to Curve V which coincides with III when $\lambda > 10$. Similarly, (162) leads to a curve which crosses III and V at $\lambda = 11$, and for all subsequent values practically coincides with IV. From a practical standpoint these four curves are sufficiently nearly alike that any one of them might be used in place of any other.

Curves I and II, however, which are obtained from formulæ (165) and (158) and therefore correspond to the first row of Table XXXI, differ from the others by amounts which are of engineering importance. For instance, when the number of sources is 15 they allow these sources to originate about 20 per cent more traffic than is allowable when the other formulæ are adopted.

In other words, little change of practical consequence is introduced in our results by shifting from the assumption of lost calls held to the assumption of lost calls cleared; or by shifting from Assumption 8 to its modified form 9. *The only difference which is of serious consequence comes from using, on*

the one hand, the assumption of independent sources (Assumption 7) and on the other the assumption that the chance of a new call being made during a short period of observation is not affected by the state of the system when that period of observation begins (Assumption 8 or 9).

§ 128. *Charts for Purposes of Computation*

Since all of the formulæ of Table XXXII group themselves into two similar classes in such a way that members of the *same* class give very similar results, while members *of different* classes do not agree so well, it will be sufficient for the further purposes of this study, as well as for most practical needs, to confine attention to a typical pair. For this purpose that pair is chosen which corresponds to the extreme conditions represented by Curves I and IV in Figs. 39 and 40. All the other formulæ give results which lie intermediate to these two but agree with the one or the other of them sufficiently well that no account need generally be taken of the differences.

Fig. 41 is a working chart computed in accordance with equation (165). The entire figure corresponds to a loss of one call per thousand. Each curve corresponds to a group of channels, the size of which is indicated by the attached number. The numbers along the left-hand margin represent the number of sources, while the numbers at the bottom give values of 3600ϵ (that is, n, the number of calls *per hour*, multiplied by T, the holding time *in seconds*[1]).

As an illustration of the use of this chart, suppose it is desired to assign to a group of ten channels a group of sources, each of which originates on the average three calls of 100 seconds' duration per hour. Then $nT = 300$. Entering the chart, it is found from the curve for $\nu = 10$ that the ordinate corresponding to this value is $\lambda = 41$. Hence 41 sources may be assigned to the group of ten channels.

As another illustration, suppose 200 sources are to be accommodated by switches capable of reaching ten trunks each. Suppose on

[1] In the telephone industry, calling rate and holding time are usually stated in this way.

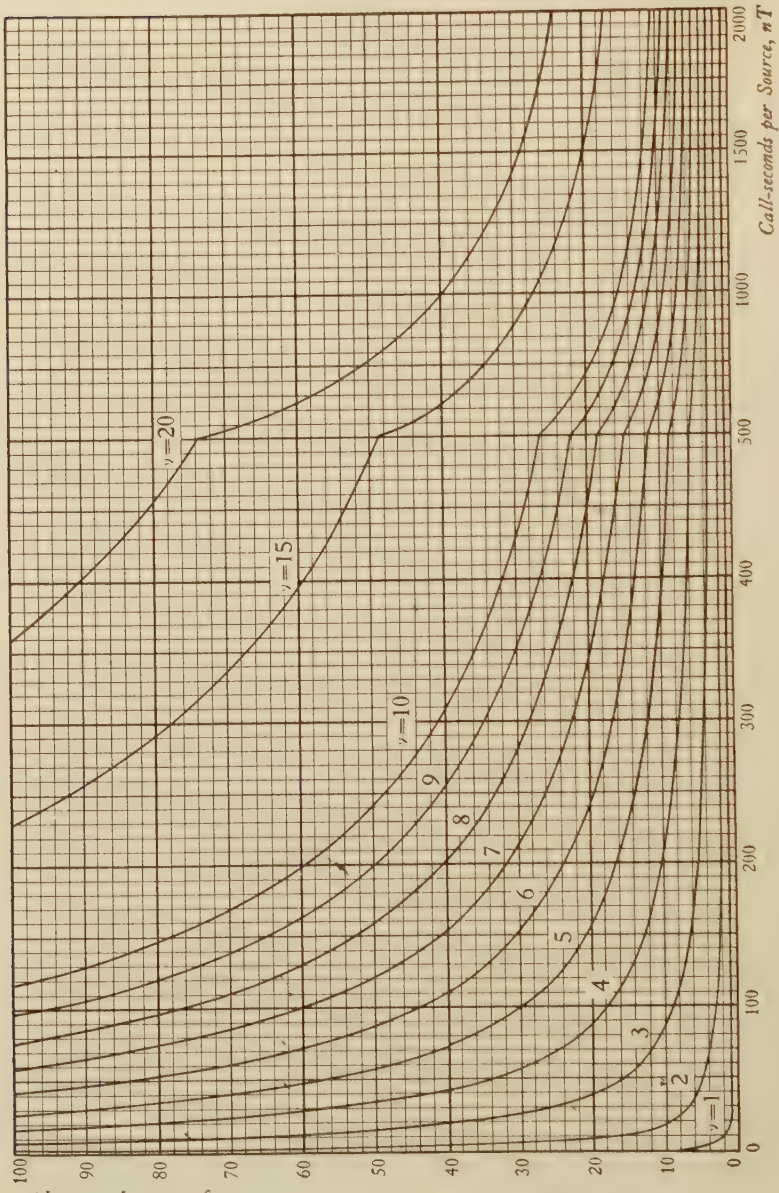


FIG. 41.—WORKING CHART FOR THE ENGSET FORMULA. PROBABILITY OF LOSS = 0.001.

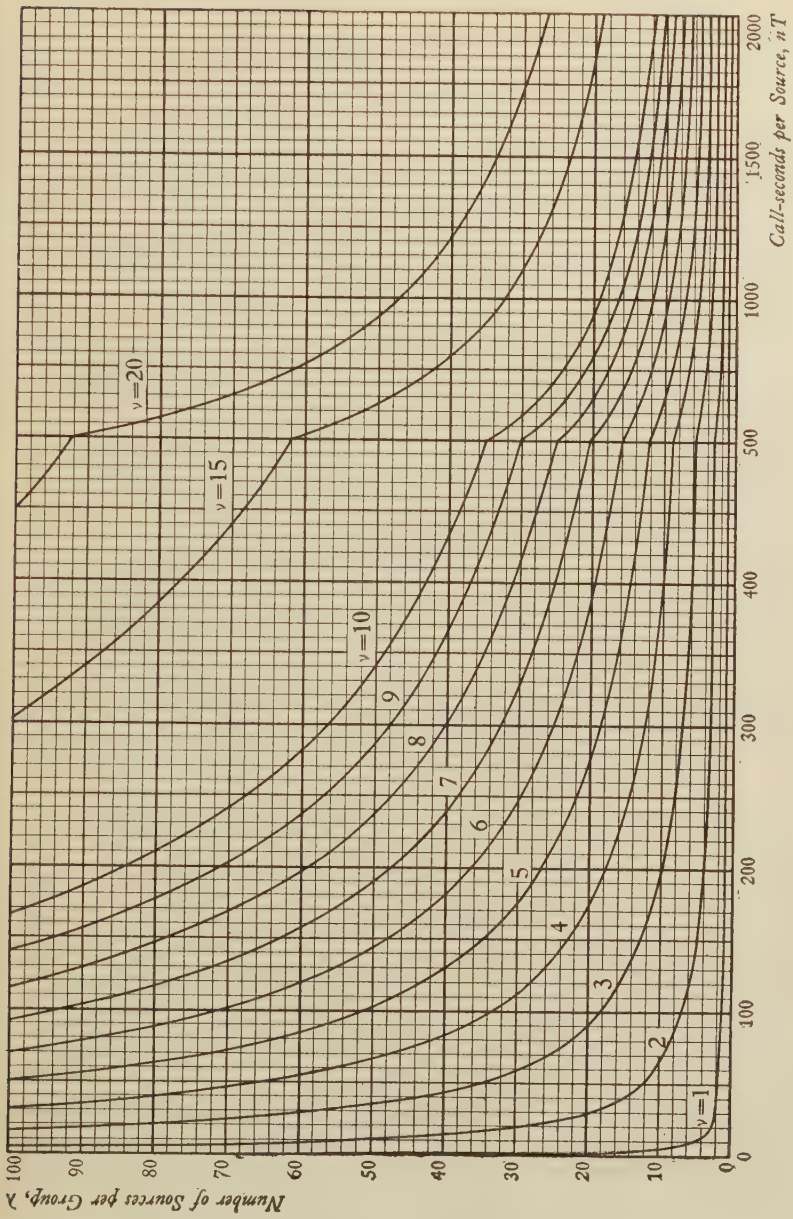


FIG. 42.—WORKING CHART FOR THE ENGSET FORMULA. PROBABILITY OF LOSS = 0.01.

the average these sources originate during the busy hour two calls of an average duration of 140 seconds, and that it is required to find how they shall be grouped. Multiplying the calling rate by the holding time gives the number 280. Entering the chart with this value it is found that each group of trunks is capable of accommodating 43 sources. Therefore 4 full groups of trunks are required. There then remain 28 sources to be accommodated by the odd group. The point upon the chart which corresponds to $\lambda = 28$ and $3600\,\epsilon = 280$ lies between the curves marked 7 and 8. Hence the odd group will require 8 channels to carry its traffic. The grouping of the sources will therefore be 4 groups of 43 and 1 group of 28, and the channels required will be 4 groups of 10 and 1 group of 8, or a total of 48.

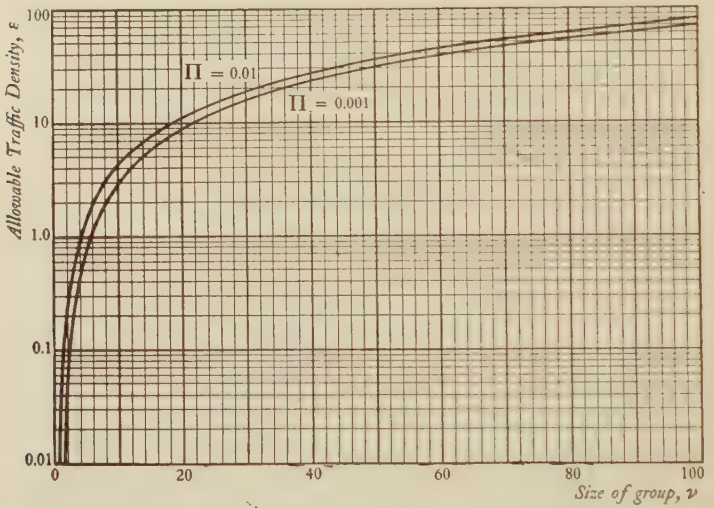


FIG. 43.—WORKING CHART FOR THE POISSON FORMULA.

Fig. 42 is a similar chart except that it corresponds to a probability of loss of one call per hundred. Its use is identical with that of Fig. 41.

In Fig. 43 are given working curves corresponding to the formula (159). Their use is slightly different from that of the curves in Figs. 41 and 42. The size of the group of channels is now represented by the numbers along the horizontal axis instead of by those on the curves, while the curves themselves correspond to a particular value of the probability of loss. Curves are given for $\Pi = 0.01$ and $\Pi = 0.001$. The

vertical axis now represents the maximum allowable traffic density, from which the number of sources must be determined since the values of λ do not explicitly occur.

The use of this chart may be illustrated by solving exactly the same problems as before. In the first case a group of 10 channels is available. Entering the chart, the allowable traffic density for such a group (for a loss of one call in a thousand) is found to be $\epsilon = 2.96$. The number of sources which can be accommodated is the largest number the traffic from which does not exceed this density. The traffic density for a single source is $nT = 0.0833$. Hence the number of sources which can safely be accommodated is $2.96/0.0833 = 35$. This number corresponds to the 41 obtained from the use of Fig. 41.

In the second illustration, where it is necessary to accommodate 200 sources originating on the average two calls of 140 seconds holding time apiece, the average traffic density of a source is $nT = 0.0778$. Since the traffic density of a group may be 2.96 we find that the number of sources which can be accommodated is $2.96/0.0778 = 38$. There are therefore required 5 full groups of 10 trunks each, together with an odd group sufficiently large to handle the traffic from the remaining 10 sources. These 10 sources give rise to a traffic density amounting to $0.0778 \times 10 = 0.778$. Entering the chart with this value it is found that the number of trunks required for this odd group is 5. The total number of trunks is therefore 55. The difference between this result and that obtained from Fig. 41 is about 14 per cent.

By means of charts such as these, computations can be carried out without the expenditure of an undue amount of time. That is, the complicated formulæ listed in Table XXXII can actually be reduced to a form in which their use for practical purposes is feasible. Figs. 41, 42 and 43 do not cover a sufficient range of values to make them satisfactory for such purposes; and, indeed, it would be difficult to print in pages such as these, charts on a scale large enough to be of much practical value. It is evident, however, that constructing such charts is possible; and since they need be made but once, the fact that the computations involved in producing them are tedious is a matter of minor importance.

PROBLEMS

1. In a certain boot-black "parlor" of a railroad terminal, the expectation of the number of customers during the peak period is 75 per hour. On the average a shine requires 4 minutes. Assume that a prospective customer, if the chairs are all full, merely looks in the door and walks away. Assume also that the customers arrive individually and collectively at random. What probability formula would you use to find the number of chairs required in order that the proportion of lost trade should not exceed 0.01?

2. Use the Poisson Formula as a means of obtaining a first approximation to the number of chairs required in the last example. Then find, by computation for neighboring values of ν, the correct number.

3. It is desired to measure the number of bursts of static per second by means of a recording chronograph. Assuming that the expected number per second is $n = 1.7$, and that the expected duration of each is $T = 0.3$ second, what proportion will not give distinct signals?

(Assume that overlapping bursts will be recorded as one; also that the bursts occur individually and collectively at random.)

4. The chronograph record consists of a number of separate entries of measurable duration. It is possible to read directly, therefore, the number of separate entries n^* and their average duration T^*. Due to overlapping, however, these are not equal to n and T. Develop formulæ for n and T in terms of n^* and T^*.

5. During an hour, such a chronograph record showed 18,241 separate entries, the aggregate length of which was 2977 seconds. What figures do you deduce for the frequency and duration of the static pulses?

§ 129. *Some Hunting Problems*

We now turn to a different class of problems: those which concern the amount of traveling which the "switch" must do in steering a call along its way. Such problems are of importance for two reasons: first, because the amount of wear to which a switch is subject ordinarily decreases with decreasing travel; second, because in many cases the time consumed in the hunting operation increases by just that much the period required to complete the connection. We choose again a

number of problems which typify, in the main, the methods of solution, without requiring an explanation of any of the technical details of telephony.

In the first place, we must notice that the "switches" themselves may be of either of two types. Each switch may be permanently connected to a "source," its function being to select a channel for that source to use when the source needs it. Such switches are technically known as "selectors." Or it may be permanently connected to the channel and, when a channel is needed by one of its group of sources, it may go in search of the "calling" source. Such switches are technically known as "finders." In either case, of course, the "group of switches" is synonymous with one of the groups

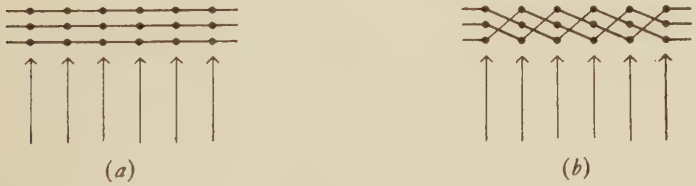


FIG. 44.

about which we have been speaking — that is, it is immaterial whether we speak of a "group of sources" or of the "group of selector switches" to which those sources are connected; and the same is true of channels and finder switches. In our present study it will be simpler to think in terms of the switches in each case.

The second point which we must notice especially, is that though the switches are assumed to be identical and to reach the same group, they may not reach the various members of that group in the same order. To be explicit about this, we may think of a group of six selectors, represented schematically by the arrows of Fig. 44; and we may suppose that they reach a group of three channels represented by the horizontal lines. So far our description applies equally well to either part (*a*) or part (*b*) of the figure. But all six switches in the arrangement (*a*) reach channel 1 first, channel 2 next, and

channel 3 last; while in arrangement (*b*) each channel appears in the lowest ("preferential") position before two switches, in the second-choice position before two others, and finally as last-choice positions for the remaining two. Condition (*a*) is known technically as a "straight multiple" and condition (*b*) as a "slipped multiple".[1]

The third point which we must notice concerns the behavior of the switch when it is released from service. Customarily it does one of two things: (*a*) returns to a "rest position," so that all idle switches are lined up in a neat schematic row as shown in Fig. 44; or (*b*) stays where it is, so that idle switches may be most anywhere, as shown in Fig. 45.

FIG. 45.

Finally, in the fourth place, the switches may hunt singly or in groups. When they hunt as a group, the switch which first succeeds in its search takes charge of the call and the rest stop hunting.

Naturally, a difference in any of these three essentials may very profoundly affect the length of hunt; so separate consideration must, in general, be given to each possibility. We shall take up a number of cases in order.

§ 130. *Individual Hunting from a Normal Position*

The simplest of all possible problems is that of a single finder hunting over a straight multiple. Obviously, how far it must go is determined solely by the position of the source which wants it. If there are ν sources and all are equally likely to be busy, the expected number of terminals tested will be

ONE FINDER STARTS FROM NORMAL REST POSITION: STRAIGHT MULTIPLE.

$$\epsilon(k) = \frac{\nu + 1}{2}; \qquad (170)$$

[1] Not all "slipped multiples" are arranged exactly as in (*b*), but we shall use the term for this simplest arrangement only.

and the probability of testing more than a specified number k will be

$$p(> k) = 1 - \frac{k}{v}. \qquad (171)$$

If the multiple, instead of being "straight," is "slipped," the formulæ (170) and (171) are still unchanged. The principal difference between this case and the former one lies in the fact that in this case all sources get equivalent grades of service, since all appear equally often in the favorable position, while in the former case those which appear nearest the rest position get a better grade of service than the others. The grade of service, averaged over all lines, is the same in both cases, however.

ONE FINDER STARTS FROM NORMAL REST POSITION. SLIPPED MULTIPLE.

For our second problem we may state the conditions as follows:

Each channel appears in the same position before every switch; all switches start from rest; if every channel is busy the switches do not repeat the test, but return to normal and discard the call. Under these circumstances it is evident that the channels are not equally used. The first to be tested will be in use most of the time, while the last in the group will seldom be busy. It is not true, however, as might be supposed at first thought, that when six channels are busy it is always the lowest six, for the following reason: If, when the last call preceding the one under consideration was made, the first six channels were busy, this last preceding call was assigned to the seventh channel; but between that time and the present, one of the six calls which were originally in progress may have been discontinued. This would leave exactly six busy channels, but they would not be the lowest six and a hunt of seven terminals would not be required. For example, if the call which occupied the lowest channel in the group has been discontinued, the switch which handles our present call need only test this lowest

ONE SELECTOR STARTS FROM NORMAL REST POSITION. STRAIGHT MULTIPLE.

channel in order to find accommodation; so instead of hunting seven terminals it need hunt one only.

What we must find is the probability that *the first k channels are busy and the next is idle.* This we may easily do by noting what would happen if, for some reason, all those calls which did not find service among these k channels were instantly cleared. Obviously this would have no effect whatever upon the service rendered by these k channels. *Hence the chance of the first k channels being simultaneously busy is equal to the probability of all k busy on the basis of formula* [1] *(167).*

Hence we have [2]:

$$ {}_{[k]}P = {}^{\infty}\Pi''_k = \frac{\frac{\epsilon^k}{k!}}{1 + \epsilon + \frac{\epsilon^2}{2!} + \cdots + \frac{\epsilon^k}{k!}}. \qquad (172) $$

The probability which we desire — that is, ${}_{[k]}P_{k+1}$ — may be obtained by subtracting from all the cases in which the first k channels are busy those cases in which the $(k + 1)$st channel is also busy. The latter cases, however, are represented by the probability ${}_{[k+1]}P$, which is given by the same law as ${}_{[k]}P$. We therefore have for the probability of hunting exactly $k + 1$ terminals

$$ p(k + 1) = {}_{[k]}P - {}_{[k+1]}P. \qquad (173) $$

Similarly the probability of hunting exactly k terminals is

$$ p(k) = {}_{[k-1]}P - {}_{[k]}P = {}^{\infty}\Pi''_{k-1} - {}^{\infty}\Pi''_k. \qquad (174) $$

---

[1] We shall assume that calls are distributed individually and collectively at random. Other assumptions could easily be used, but this will be quite sufficient for our present purposes.

[2] The notation has this significance: ${}_1P$ means " the first trunk busy "; ${}_{1,3}P_2$ means " the first and third busy and the second idle "; ${}_{[4]}P_5$ means " the first four busy and the fifth idle "; ${}_6P_{[5]}$ means " the first five idle and the sixth busy. The scheme, I think, is obvious.

The expected hunt under these circumstances is [1]

$$\epsilon(k) = \sum_{k=1}^{\nu} k\, p(k).$$

This can be thrown into the form

$$\epsilon(k) = \sum_{k=0}^{\nu-1} {}_{[k]}P, \tag{175}$$

which is more suitable for numerical calculation.

If the multiple is so arranged that each channel appears as often in one position as in any other the situation is quite different. In this case busy channels will be distributed in haphazard fashion over the entire bank instead of being concentrated near the bottom, and the chance of an idle channel being near the bottom of the group will be materially greater than in the preceding example. The solution of the problem is obtained by the following argument:

ONE TRUNK SELECTOR STARTS FROM NORMAL REST POSITION. SLIPPED MULTIPLE.

The probability of exactly j busy channels is denoted by [2] $P(j)$. If these j busy channels are distributed at random over the entire group, the chance that the first k tested by the switch are all busy is

$${}_{[k]}P = \frac{C_k^j}{C_k^\nu}. \tag{176}$$

The probability of testing more than k terminals is found by summing the product of these two expressions for every possible value of j. Formally it is given by the formula

$$p(> k) = \sum_{j=k}^{\nu} {}_{[k]}P \cdot P(j).$$

[1] Note how the assumption of "lost calls cleared" is justified by the fact that, when all ν channels are busy, the switches discard the call. Hence the chance of hunting just ν terminals is just ${}_{[\nu-1]}P$, while $p(> \nu) = 0$.

[2] We shall use for it the formula ${}^{\infty}P''(j)$, to conform to the conditions as we have laid them down. In other circumstances some other formula might be needed.

Using formula (166) for $P(j)$ we find that

$$p(>k) = \frac{(\nu-k)!}{\nu! \sum_{i=0}^{\nu} \frac{\epsilon^i}{i!}} \sum_{j=k}^{\nu} \frac{\epsilon^j}{(j-k)!}.$$

If, now, we call $j-k$ by a new symbol (which may as well be i as anything, since the two summations are entirely independent) this becomes [1]

$$p(>k) = \epsilon^k \frac{(\nu-k)!}{\nu!} \frac{\sum_{i=0}^{\nu-k} \frac{\epsilon^i}{i!}}{\sum_{i=0}^{\nu} \frac{\epsilon^i}{i!}} = \frac{{}^{\infty}\Pi_{\nu}''}{{}^{\infty}\Pi_{\nu-k}''}. \tag{177}$$

As before,

$$p(k) = p(>k-1) - p(>k). \tag{178}$$

We may also find the expected hunt, which is

$$\epsilon(k) = \sum_{k=0}^{\nu} k\, p(k).$$

This, however, can be further reduced to a form which is more suitable for purposes of computation by noting that when written out in full it is:

$$\begin{aligned}\epsilon(k) = p(>0) &- p(>1)\\ &+2\,p(>1) - 2\,p(>2)\\ &\qquad +3\,p(>2) - \dots\\ &\qquad\qquad \dots -(\nu-1)\,p(>\nu-1)\\ &\qquad\qquad\quad +\nu\,p(>\nu-1),\end{aligned}$$

which is obviously equal to

$$\epsilon(k) = \sum_{k=0}^{\nu-1} p(>k). \tag{179}$$

---

[1] Except for $k=\nu$. It is obvious, from a common-sense standpoint, that the switch cannot hunt more than ν terminals. Hence $p(>\nu)=0$. The formula fails to give this because the switch might (logically) fail in its νth trial if all channels were busy, and in that case a $(\nu+t)$th trial *would be needed* for success.

As an illustration of the extent to which the average hunt is reduced by slipping the multiple, the results of a numerical example computed in accordance with each of these formulæ may be presented. The case chosen is one in which the total number of channels is $n = 100$ and the expected number busy is $\epsilon = 34.49$. If each channel appears in the same position before all the switches the average number of terminals tested is found to be 19.6, while if the multiple is slipped the average number of tests is only 1.53.

§ 131. *Individual Hunting with Stay-Put Switches*

In studying the problem of individual hunting with stay-put switches we shall assume that the switch is capable of hunting over the entire group of channels no matter from what position it may start, but that if no idle trunk is then found it will not repeat the test. Otherwise when all trunks are busy the number of terminals tested before an idle trunk is found would depend upon the length of time which elapses before a trunk becomes idle and the speed with which the test is made, both of which questions we wish to avoid.

We consider only the case of selector switches, as the other case is trivial. It is obvious that after the system has been in operation for a certain length of time, the switches will be distributed at random over the terminal bank. Since the busy channels are likewise distributed at random over the bank they are also at random with respect to any switch. Hence in this case the hunting probabilities follow exactly the same law as if the switches started from a normal position and hunted over a slipped bank. The formulæ to be applied are therefore (177) and (178).

ONE STAY-PUT SELECTOR HUNTING OVER A STRAIGHT MULTIPLE.

Since the positions of the busy channels are distributed at random with respect to any one switch, even when the multiple is straight, it follows that no change is introduced if a slipped multiple is used, and formulæ (177) and (178) again apply.

ONE STAY-PUT SELECTOR HUNTING OVER A SLIPPED MULTIPLE.

§ 132. *Group Hunting with Stay-Put Switches*

Under this heading we shall consider two distinct cases distinguished in the following manner: When a group of switches of the stay-put variety are all started at once it may happen that the first to reach the desired terminal [1] may be, not a single switch, but two or more which are accidentally moving together. The problem then arises as to what disposition is to be made of them, for it is obviously not desirable to allow them all to connect with it. There are two alternatives: one is to allow some one of the group to seize the terminal; the other is to pass them all by and wait until it is tested by a switch which is travelling alone.

In the first case, if λ switches are searching over a straight multiple, the solution is easily obtained by this line of thought: Any individual switch is just as likely to be on one terminal as another, quite independently of where the other switches may be. Then let the heavy line of Fig. 46 be the calling source, there being in all ν sources and λ switches. The chance that a particular switch is on some one of the k bracketed sources to begin with is k/ν; and the chance that it is *not* there is $1 - k/\nu$. As the positions of the switches are quite independent, the chance that no one of the λ switches is on any of the bracketed sources is $(1 - k/\nu)^{\lambda}$. But this is just the condition under which the switches would have to make *more than* k tests in order to reach the calling source. Hence we have:

GROUP OF STAY-PUT FINDER SWITCHES HUNTING OVER A STRAIGHT MULTIPLE. IF TWO ARRIVE AT ONCE, ONE TAKES CHARGE OF THE CALL.

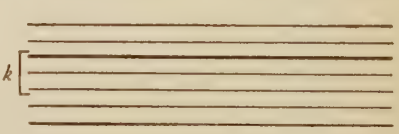


FIG. 46.

$$p(> k) = \left(1 - \frac{k}{\nu}\right)^{\lambda}. \qquad (180)$$

[1] "Terminal" is here used as a general term meaning "source or channel."

As before

$$p(k) = p(>k-1) - p(>k) = \left(1 - \frac{k-1}{\nu}\right)^{\lambda} - \left(1 - \frac{k}{\nu}\right)^{\lambda}. \quad (181)$$

The expectation of k — that is, the number of tests we may expect them to make before giving service — is

$$\epsilon_1(k) = \Sigma\, k\, p(k) = \frac{1^{\lambda} + 2^{\lambda} + 3^{\lambda} + \ldots + \nu^{\lambda}}{\nu^{\lambda}}. \quad (182)$$

In the second case which we have mentioned the formula is decidedly more complicated, and requires the use of the principle of alternative compound probabilities for its evaluation. We begin as follows:

GROUP OF STAY-PUT FINDER SWITCHES HUNTING OVER A STRAIGHT MULTIPLE. IF TWO OR MORE ARRIVE AT ONCE NEITHER TAKES CHARGE OF THE CALL.

If there are just i' switches resting on the k bracketed sources of Fig. 46, the chance that there are exactly i on the source next below them is given by the formula

$${}_k p_{i'}(i) = C_i^{\lambda - i'} \left(\frac{1}{\nu - k}\right)^{i} \left(1 - \frac{1}{\nu - k}\right)^{\lambda - i' - i}.$$

Evidently if the i' switches which rest on the k sources are so arranged that no source has exactly one switch, the test must exceed k terminals. That is, it will be more than $k + 1$ if $i \neq 1$, while if $i = 1$ it will be exactly $k + 1$. Hence if we knew the probability of there being i' switches on the bracketed sources so arranged that no source has on it exactly one switch, we would be able to compute the probability of a hunt of any desired magnitude by merely building up the proper form of alternative compound probability. For the moment we may content ourselves with writing a symbol for it, in the hope that later on we may be able to find a formula for it. Let us choose for this purpose the notation ${}_k P(i')$, the prefixed subscript being added in this case, as in the case of ${}_k p_{i'}(i)$, to call attention to the fact that the probability depends upon the

particular value of k which we choose to consider. Then we have at once, as a formal expression for our solution,

$$p(k+1) = \sum_{i'=0}^{\lambda} {}_{1}P(i') \; {}_{k}p_{i'}(i). \tag{183}$$

As for the determination of an expression for ${}_{k}P(i')$, we notice that if there are exactly i' switches on $k+1$ sources, and if these switches are arranged in the fashion described, it must be true, either that all are on the first k and none on the $(k+1)$st, or else all but *two* are on the first k and those two are on the $(k+1)$st, or else all but three are on the first k and those three on the $(k+1)$st, or some similar arrangement. The only cases which are excluded are those which would require only one on the first k or only one on the $(k+1)$st, since either of these cases would require a *single* switch on some source. Hence we have the recursion formula

$$\begin{aligned} {}_{k+1}P(i') &= {}_{k}P(i') \; {}_{k}p_{i'}(0) + {}_{k}P(i'-2) \; {}_{k}p_{i'-2}(2) \\ &\quad + {}_{k}P(i'-3) \; {}_{k}p_{i'-3}(3) + \cdots \\ &= \sum_{i=0}^{i'} {}_{k}P(i'-i) \; {}_{k}p_{i'-i}(i), \end{aligned}$$

it being understood that the values $i=1$ and $i = i'-1$ are not to be included in the summation. From this recursion formula it is possible to obtain ${}_{k+1}P(i')$ if ${}_{k}P(i')$ is known. Hence if the values of this function are known for any one value of k they may be found for all others. But it is obvious at once that

$${}_{1}P(i') = C_{i'}^{\lambda} \frac{(\nu-1)^{\lambda-i'}}{\nu^{\lambda}}.$$

With this as a start, it requires only routine algebra to show that[1]

$${}_{2}P(i') = C_{i'}^{\lambda} \frac{(\nu-2)^{\lambda-i'}}{\nu^{\lambda}} (2^{i'} - 2i' + i' \, C_{i'-2}^{0}),$$

$${}_{3}P(i') = C_{i'}^{\lambda} \frac{(\nu-3)^{\lambda-i'}}{\nu^{\lambda}} [3^{i'} - 3i' \, 2^{i'-1} + 3i'(i'-1) - i'(i'-1) \, C_{i'-3}^{0}],$$

[1] These expressions have different values according as i' is, or is not, equal to k. This has been taken account of by the terms having the factors $C_{i'-k}^{0}$ which vanish for all values of i' except $i' = k$.

and in general

$$_kP(i') = C^{\lambda}_{i'}\frac{(\nu-k)^{\lambda-i'}}{\nu^{\lambda}}\left(\sum_{h=0}^{k-1}(-1)^h C^k_h \frac{i'!}{(i'-h)!}(k-h)^{i'-h} + (-1)^k \frac{i'!}{(i'-k+1)!}C^0_{i'-k}\right).$$

We now have formulæ for both $_kP(i')$ and $_kp_{i'}(i)$, and can therefore substitute them in (183). Some more routine algebra then shows that

$$p(k+1) = \sum_{h=0}^{k}(-1)^h \frac{\lambda!}{(\lambda-h-1)!}C^k_h \frac{(\nu-h-1)^{\lambda-h-1}}{\nu^{\lambda}} \qquad (184)$$

when $k+1 < \nu$, while

$$p(\nu) = \sum_{h=0}^{\nu-1}(-1)^h \frac{\lambda!}{(\lambda-h)!}C^{\nu-1}_h \frac{(\nu-h)^{\lambda-h}}{\nu^{\lambda}}. \qquad (184)$$

For completeness we quote also the expected number of tests:

$$\epsilon_1(k) = \frac{\lambda!}{\nu^{\lambda}}\sum_{h=0}^{\nu-1}(-1)^h C^{\nu}_{h+1}\frac{(\nu-h)^{\lambda-h}}{(\lambda-h)!}. \qquad (185)$$

These formulæ probably serve no other useful purpose, so far as this text is concerned, than that of showing how complicated it is possible for problems of this general type to become.

The method by means of which (180) and (181) are derived can be applied without change to the case of a slipped multiple. Hence these formulæ are equally valid in this case. The same is true of (184) and (185). In both cases the mental picture upon which the argument is based requires some modification; but the steps to be followed and the results themselves are identical throughout.

GROUP OF STAY-PUT FINDER SWITCHES HUNTING OVER A SLIPPED MULTIPLE. IF TWO OR MORE ARRIVE AT ONCE, ONE TAKES CHARGE OF THE CALL.

GROUP OF STAY-PUT FINDERS HUNTING OVER A SLIPPED MULTIPLE. IF TWO OR MORE ARRIVE AT ONCE, NEITHER TAKES CHARGE OF THE CALL.

If instead of finders, however, we are dealing with selectors,

the hunts are in all cases likely to be shorter. If two or more can arrive at once and still take charge of the call, the formula by means of which the results are expressed is not very difficult to obtain. We fall back upon the similar case with only one selector, for which the solution is given by (177) and (178). It is obvious that any member of the group at present under consideration might be the *one selector* previously considered. Hence each of the group must obey the laws obeyed by that one. This being understood, the chance that some i switches would need to hunt exactly k terminals while the rest hunt more than k is

GROUP OF STAY-PUT SELECTORS HUNTING OVER A STRAIGHT MULTIPLE. IF TWO OR MORE ARRIVE AT ONCE, ONE TAKES CHARGE OF THE CALL.

$$C_i^\lambda [p_1(k)]^i [p_1(>k)]^{\lambda-i},$$

the $p_1(k)$ being given by (178) and $p_1(>k)$ by (177). Hence the chance of a hunt of just k terminals is [1]

$$\begin{aligned} p_\lambda(k) &= \sum_{i=1}^{\lambda} C_i^\lambda [p_1(k)]^i [p_1(>k)]^{\lambda-i} \\ &= [p_1(k) + p_1(>k)]^\lambda - [p(>k)]^\lambda \\ &= [p_1(>k-1)]^\lambda - [p_1(>k)]^\lambda \\ &= \left(\frac{{}^{\infty}\Pi_\nu''}{{}^{\infty}\Pi_{\nu-k+1}''}\right)^\lambda - \left(\frac{{}^{\infty}\Pi_\nu''}{{}^{\infty}\Pi_{\nu-k}''}\right)^\lambda. \end{aligned} \tag{186}$$

The chance of a hunt of more than k terminals is

$$p_\lambda(>k) = [p_1(>k)]^\lambda = \left(\frac{{}^{\infty}\Pi_\nu''}{{}^{\infty}\Pi_{\nu-k}''}\right)^\lambda, \tag{187}$$

and the expected hunt is

$$\epsilon_1(k) = \sum_{k=0}^{\nu-1} \left(\frac{{}^{\infty}\Pi_\nu''}{{}^{\infty}\Pi_{\nu-k}''}\right)^\lambda. \tag{188}$$

As an illustration of the extent to which group hunting may reduce the average hunt we may consider the same illustration as before.

---

[1] Except for $k = \nu$. The formula in this case lacks the negative term.

We found that with a straight multiple and a switch which starts from normal, the average number of terminals tested was 19.6. With stay-put switches, only one being assigned, or with a random slip in the multiple, this was reduced to 1.53. If the switches stay put and are started in groups of two, three, four or five, the answers are, respectively, 1.14, 1.04, 1.014 and 1.005. Remembering that the minimum is a test of one terminal, the extent to which the test is reduced by starting a group is evident.

§ 133. *The Problem of Double Connections*

One type of problem which frequently presents itself in connection with the use of apparatus by a considerable number of different people is that of preventing one person from seizing what is already in use and thus causing inconvenience to somebody else. This is frequently accomplished by operating a relay, or some similar device, which cuts off access to the particular channel which has been assigned. Obviously, such a blocking device will ordinarily require time for its operation, and during a portion of this time at least it will be possible for another source to seize the already busy channel, thus creating what is technically termed a "double connection." We are interested in determining what proportion of calls can be expected to suffer inconvenience from this source.

Problems of this sort ordinarily arise in checking up whether a proposed system does or does not meet certain specified standards, and since we can tolerate much larger errors in studying such problems than in studying losses, we need not be so careful of the exactness of our assumptions. In particular, we may assume that the calls occur individually and collectively at random, and that, in assigning channels to calling sources, it is a matter of pure chance which of the idle channels is chosen. As always, we denote the number of channels by ν, the calling rate by n, and the expected holding time by T.

Let us suppose that we observe the group for a short time

dt, at the beginning of which exactly j channels are busy. The chance that a call is made during this interval is obviously $n\lambda dt$; and if so, the chance that it is assigned to some *particular* channel which we have set out to watch is $n\lambda dt/(\nu - j)$.

Next we write down, by the use of alternative compound probabilities, the chance $p(b)$ that this particular channel is seized during dt, if we do not know how many channels are already in use. We assume that the number of calls originated per unit of time is n and that the number of channels is ν. Denoting by $P(j)$ the probability that exactly j channels are busy at the beginning of this interval, it is:

$$p(b) = \sum_{j=0}^{\nu-1} P(j) \frac{n\lambda dt}{\nu - j} = n\lambda dt \sum_{j=0}^{\nu-1} \frac{P(j)}{\nu - j}. \tag{189}$$

If we assume that lost calls are instantly cleared, $P(j)$ must be given the form (166); whence (189) becomes

$$\frac{dt}{T} \frac{\sum_{j=0}^{\nu-1} \frac{\epsilon^{j+1}}{(\nu - j)\, j!}}{\sum_{j=0}^{\nu} \frac{\epsilon^{j}}{j!}}, \tag{190}$$

which we shall denote simply as $f(\epsilon)\, dt/T$, $f(\epsilon)$ meaning, of course, the fraction in (190).

It remains to determine the relation of this result to the probability of double connections. From the fashion in which we have derived it we are assured that it represents the probability of a call being made upon a particular idle channel during a particular short interval of observation. One way of making this observation, however, would be to introduce a call and see if it is interfered with. Hence it is obvious that what we have obtained is the probability that our call will be followed by another within so short a time as to result in a double connection. In other words $f(\epsilon)\, dt/T$ is the probability that a call will be involved in a double connection by virtue of the fact that another call *originating later than itself* obtains access to the same trunk. It is just as likely, however, to interfere with someone else who arrived earlier, as to be

interfered with by someone who arrives later. Hence $f(\epsilon)dt/T$ is only half of the total probability that a call will be involved in a double connection. This leads us at once to

$$p(dc) = 2p(b) = \frac{2\,dt}{T} f(\epsilon). \qquad (191)$$

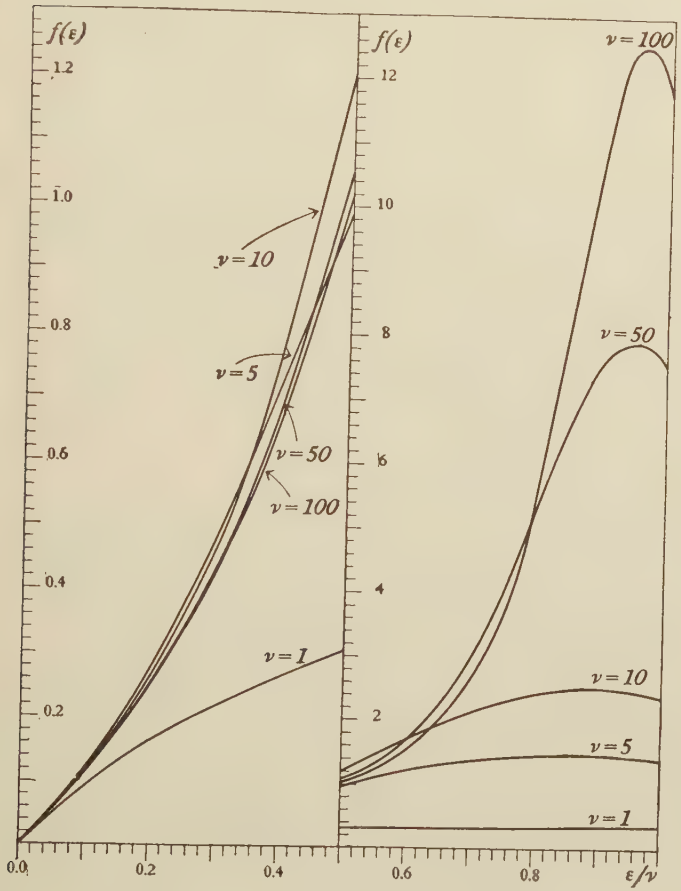


FIG. 47.

In order to facilitate computations, a chart of the function $f(\epsilon)$ has been prepared for a considerable number of values of ν between 1 and 100. This chart, with the abscissæ measured in terms of ϵ/ν instead of ϵ for the sake of compactness, is

presented as Fig. 47. Its use may be illustrated by a simple example:

Suppose there are 20 trunks handling 288 calls per hour, each with an average holding time of 100 seconds. Then $T = 100$, $n\lambda = \frac{288}{3600} = 0.08$, $\epsilon = n\lambda T = 8$ and $\nu = 20$; so that $\epsilon/\nu = 0.04$.[1] Referring to Fig. 47, we find that for this case $f(\epsilon) = 0.720$. Suppose that the unguarded interval is 0.05 second, then the probability of a double connection is:

$$p(dc) = \frac{2 \times 0.05}{100} 0.720 = 0.00072.$$

This means that under these circumstances, 72 calls out of every 100,000 would be involved in double connections, due account being taken of the fact that every double connection involves two calls.

Other problems involving double connections can be worked equally well provided the value of $f(\epsilon)$ is taken from the proper one of the curves of Fig. 47.

§ 134. *Delays in Awaiting Service*

The problems presented by systems which operate upon a delay basis instead of a loss basis — that is, in which a call is not discarded when there is no apparatus to handle it, but is merely held over until something becomes free — are much the most complicated with which the traffic statistician must deal. There are several reasons for this:

In the first place, we have seen in our study of the probability of loss that our results are quite independent of the lengths of individual calls. They are the same in any two systems which possess the same traffic density, whether that traffic density be made up of many short calls or a few long ones, and whether or not the calls are all of the same length, or of different lengths. When we come to the study of delay problems, however, this is no longer true. It can be seen at a glance, for example, that if the calls are all of like length, the delays will be greater the greater that length may be.

[1] These values are given with greater accuracy than is obtainable from the sketchy diagram of Fig. 47.

Thus, a system which had an expectation of three twenty-minute calls per hour, and one which had an expectation of sixty one-minute calls, should be identical so far as probability of loss is concerned; but it is obvious from a common-sense standpoint that the person who sought service and found no available apparatus in the first case would face the prospect of a longer wait before some became idle than in the second. It is perhaps not so easy to see from an intuitive standpoint that the distribution of call lengths — whether they are all of like length or not, and if not, how they differ — also affects the problem; but once we come to formulating our results we find that this is true.

In the second place, when we were dealing with loss problems, we were able to affirm that the end-points of calls, like their points of origin, were distributed at random. But this is no longer true in the case of systems with waiting arrangements. Thus, if calls are all of unit length, and if there are just five trunks, there can only be five calls in progress at any time, and hence no more than five can terminate within the same unit of time. They may still originate at random, but the delays to which some of them are subjected smooth out the distribution of their end-points. It is to this fact, indeed, that most of the mathematical difficulties are due.

In the third place, in a system which operates upon a loss basis, every subscriber who is inconvenienced at all is inconvenienced in just the same way as every other: his call is discarded. But in a delay system some suffer delays of negligible length, others long delays. It is no longer sufficient to specify the standard of service by a mere statement that such and such a proportion of calls is delayed: it becomes necessary to say instead what proportion is delayed *more than a specified time*. This adds one more complication to the problem.

As a result of these various complexities there are only two delay problems which have sufficiently simple solutions to justify their presentation in a text such as this. Indeed, the rest have in many instances been dealt with in an approximate

fashion only, and can be said to have been "solved" only in the sense that engineering design can be carried out with reasonable assurance that a sufficient factor of safety exists, but not in the sense that the factor of safety is known. We present these two solutions only, as illustrations of the simpler methods of attack.

In the first place, if all the calls must be accommodated by a single channel instead of by a group of channels, and if they are also all of the same length, a rigorous solution can be found. We speak of this as the case of "non-cooperative channels."

In the second place, if call lengths are governed by an exponential distribution function (which has the effect of rendering their end-points "random" in spite of the attempts of the system to smooth them out) a general solution can be obtained even if there is a group of channels instead of a single one.

§ 135. *Calls of Equal Length at Non-Cooperative Channels; The Probability of Congestion j*

We shall assume that calls originate individually and collectively at random, and that if at any time the congestion is so great that more than one is awaiting service, they will be served in the order in which they originated. As in the case of the problem of loss, we denote by $P(j)$ the chance that there are just j sources either seeking service or being served at the same time — that is, that the congestion is j. We also assume, as always, that the system is in statistical equilibrium, so that these probabilities are independent of time.

Let n be the calling rate and T the holding time. Then, since it is impossible for calls to overlap, the proportion of time during which our channel can be expected to be busy is nT, and its expected idle time is $1 - nT$. This latter, however, is obviously $P(0)$. Hence, using the notation $nT = \epsilon$ as before, we have

$$P(0) = 1 - \epsilon.$$

Next we consider an interval of length exactly equal to the holding time, and ask, What is the probability that the congestion is exactly j at the end of this interval? Obviously, there will be j calls in progress at the end of the interval provided there was none in progress when it began and just j came in meanwhile, or provided there was a congestion of 1 when the interval started and j came in meanwhile (for the one which was in progress will have discontinued before the interval ends), or provided there were 2 in progress at the beginning of the interval and $j - 1$ came in, and so on. As the system is in statistical equilibrium, the probability of any one of these states at the beginning of the interval is equal to the probability of the same state at its end, hence we arrive at the law

$$P(j) = [P(0) + P(1)]\frac{\epsilon^j e^{-\epsilon}}{j!} + P(2)\frac{\epsilon^{j-1} e^{-\epsilon}}{(j-1)!} + P(3)\frac{\epsilon^{j-2} e^{-\epsilon}}{(j-2)!} + \ldots + P(j+1)\, e^{-\epsilon}. \tag{192}$$

Suppose, now, that we write down the first few of these equations. They are:

$$P(0) = [P(0) + P(1)]\, e^{-\epsilon},$$

$$P(1) = [P(0) + P(1)]\, \epsilon\, e^{-\epsilon} + P(2)\, e^{-\epsilon},$$

$$P(2) = [P(0) + P(1)]\frac{\epsilon^2}{2!} e^{-\epsilon} + P(2)\, \epsilon\, e^{-\epsilon} + P(3)\, e^{-\epsilon}.$$

These equations can easily be solved in order, and give (when we remember that $P(0) = 1 - \epsilon$)

$$P(1) = (1 - \epsilon)(e^{\epsilon} - 1),$$

$$P(2) = (1 - \epsilon)[e^{2\epsilon} - e^{\epsilon}(1 + \epsilon)],$$

$$P(3) = (1 - \epsilon)\left[e^{3\epsilon} - e^{2\epsilon}(1 + 2\epsilon) + e^{\epsilon}\left(\epsilon + \frac{\epsilon^2}{2}\right)\right];$$

the general rule being, as can be shown by more elaborate methods, that $P(j)$ is always the product of two factors, of which the first is $(1 - \epsilon)$, and the second is composed of a

series of exponential terms with suitable coefficients. These coefficients are, in every instance after the first, the sum of two terms taken from the power series expansion of the exponential which they multiply: and specifically they are the first and second, the second and third, the third and fourth, and so on, respectively. The general formula therefore is

$$P(j) = (1-\epsilon)\sum_{k=1}^{j}(-1)^{j-k}e^{k\epsilon}\left(\frac{(k\epsilon)^{j-k}}{(j-k)!}+\frac{(k\epsilon)^{j-k-1}}{(j-k-1)!}\right). \quad (193)$$

§ 136. *Calls of Equal Length at Non-Cooperative Channels; The Expected Delay*

Our next objective is the determination of the expected delay $\epsilon(\tau)$. It is at once obvious that the sum $0\cdot P(0) + 1\cdot P(1) + 2\cdot P(2) + \ldots$ represents the aggregate expected length of all calls in a typical unit of time. This means, not the aggregate of the times during which they are obtaining service, but the sum obtained by adding to those "intervals of use," the delays as well. But the aggregate of all the intervals of use is ϵ. Similarly, the aggregate of all the delays is obtained by multiplying the expected number of calls by the expected delay $\epsilon(\tau)$. Thus we arrive at the relation

$$\epsilon + n\epsilon(\tau) = \sum_{j=0}^{\infty} j\,P(j). \quad (194)$$

From this equation the expected delay $\epsilon(\tau)$ may be obtained, provided we can evaluate the summation which occurs in the right-hand member. To do this we return to the equation (192), which for our present purpose may be written in the form

$$P(j) = \sum_{k=0}^{j} P(j+1-k)\,\frac{\epsilon^k e^{-\epsilon}}{k!} + P(0)\,\frac{\epsilon^j e^{-\epsilon}}{j!}.$$

It is impossible to substitute this formula in (194) and thereby evaluate $\epsilon(\tau)$, for the attempt to do so leads us to a worthless identity. However we can accomplish our purpose by a somewhat indirect artifice.

If we form, not the sum $\Sigma\, j\,P(j)$, but the sum $\Sigma j^2\,P(j)$

instead, and if we then interchange the order of the k and j summations in the usual manner, we arrive at the equation

$$\sum_{j=0}^{\infty} j^2 P(j) = \sum_{k=0}^{\infty} \sum_{j=k}^{\infty} j^2 P(j+1-k) \frac{\epsilon^k e^{-\epsilon}}{k!} + P(0) \sum_{j=0}^{\infty} j^2 \frac{\epsilon^j e^{-\epsilon}}{j!}. \quad (195)$$

Direct evaluation gives us the result

$$\sum_{j=0}^{\infty} j^2 \frac{\epsilon^j e^{-\epsilon}}{j!} = \epsilon(\epsilon + 1). \quad (196)$$

This takes care of the last term of (195).

As for the double summation, it can be simplified by replacing the element of summation j by a new element $h = j - k + 1$. Since $j^2 = h^2 + 2h(k-1) + (k-1)^2$, the double summation splits up into three terms:

$$\sum_{k=0}^{\infty} \frac{\epsilon^k e^{-\epsilon}}{k!} \sum_{h=1}^{\infty} h^2 P(h),$$

$$2 \cdot \sum_{k=0}^{\infty} (k-1) \frac{\epsilon^k e^{-\epsilon}}{k!} \sum_{h=1}^{\infty} h P(h),$$

$$\sum_{k=0}^{\infty} (k-1)^2 \frac{\epsilon^k e^{-\epsilon}}{k!} \sum_{h=1}^{\infty} P(h).$$

In the form in which the terms now appear the h- and k-summations are independent of one another and their values can readily be found. Actual evaluation gives for the k-summations the results 1, $\epsilon - 1$ and $\epsilon^2 - \epsilon + 1$, respectively; while the first h-summation is identical with the left-hand side of (195),[1] and the last h-summation is $1 - P(0)$. Hence, when we substitute all these relations, together with (196), in (195), we obtain

$$\sum_{k=0}^{\infty} k P(k) = \frac{\epsilon^2 - 2\epsilon}{2(\epsilon - 1)}. \quad (197)$$

We now return to (194) and note that, since $nT = \epsilon$, the

[1] The difference in the lower limit of summation is unimportant, since the term corresponding to $j = 0$ is itself zero.

left-hand member can be written $\left(1 + \frac{\epsilon(\tau)}{T}\right)\epsilon$. Combining (194) and (197) and making use of this relationship, we arrive finally at a formula for the expected delay

$$\frac{\epsilon(\tau)}{T} = \frac{1}{2}\cdot\frac{\epsilon}{1-\epsilon}. \tag{198}$$

§ 137. *Exponential Distribution of Holding Times; Delays at Cooperative Groups of Channels*

We found in our discussion of probability of loss — and the same applies equally here — that the entire problem could be expressed in terms of two "elementary probabilities": the probability that a new call arrives during a test interval dt at the beginning of which the congestion is j, and the probability of one ending during such an interval. As we are assuming that the sources which originate calls have no knowledge of the state of the system, and therefore cannot be influenced by it until they have actually placed a call, the first of these elementary probabilities is here just what it was in the problem of loss. Every difference that exists between the two types of service must therefore be attributable to some difference in the second elementary probability. Let us, then, think for a moment about the chance of a call, known to be in progress at the time $t = 0$ at which we begin to observe it, ending before $t = dt$.

Our problem contemplates no change in the nature of the call *after service is given*. Hence the fact that we observe a call to be in progress means merely that it must end within a time equal to its own holding time, and the probability that it will end during the time dt is therefore dt/T, just as before, *unless we are given information which leads us to infer something about the time at which it began*. But when the knowledge of the degree of congestion is known, this probability is altered — for otherwise, as we have already said, both elementary probabilities, and therefore $P(j)$ itself, would be the same as in the loss problem. Hence *a knowledge of the degree of con-*

gestion must, by inference at least, convey information regarding the times at which calls were given service.[1]

Now it is a peculiar property of the exponential distribution of holding times that even *absolute* knowledge of the time at which a call was given service does not affect its chance of ending during dt; and obviously if *absolute* knowledge does not, *inferential* knowledge cannot. This is why the delay problem can be easily solved when such a distribution of call lengths is postulated.

To show that the length of time a call has been in progress does not affect the probability of its termination, when the call lengths are distributed in accordance with the law [2]

$$p(T) = \frac{1}{\overline{T}} e^{-T/\overline{T}}, \tag{199}$$

is a very simple mathematical problem. Suppose the call is known to have been in progress just t seconds. It must then be at least t seconds long, the probability of which is

$$p(> t) = \int_t^\infty p(T)\, dT = e^{-t/\overline{T}}.$$

If it ends during dt it must have had a length lying between t and $t + dt$, the probability of which is

$$p(t)\, dt = \frac{dt}{\overline{T}} e^{-t/\overline{T}}. \tag{200}$$

The quotient of these [3] is the chance that the call ends during

[1] This statement is undoubtedly true, as the above argument shows, and that without regard to how the lengths of the calls may be distributed. I have frequently attempted to formulate a direct argument to replace the *reductio ad absurdum* here given. Such an argument should give a value for the elementary probability of termination under congestion j, and it would be a very easy matter to formulate a complete solution if this were known. The direct relationship, however, still remains as baffling as ever.

[2] For the time being we write $\overline{T}$ and $\overline{\tau}$, instead of $\epsilon(T)$ and $\epsilon(\tau)$, in order to simplify the appearance of our equations.

[3] Obviously $p(t)\, dt$ (the chance of a call ending during dt) is the chance of it lasting until $t = 0$ [which is $p(> t)$] multiplied by the conditional chance that it does not last beyond $t + dt$ [which we may write $p_{>t}(< t + dt)$]. The first two have been computed, and the third is exactly the probability for which we are seeking.

dt, and turns out to be $dt/\bar{T}$. As this is independent of t, the statement is proved.

§ 138. *Exponential Distribution of Holding Times; The Probability of Congestion j*

Suppose there are ν channels serving λ sources, each of which latter originates an average of n calls per unit time. Suppose, further, that the call lengths are distributed according to the exponential law (199). Then the principle of statistical equilibrium leads to the set of equations

$$\left.\begin{aligned}
&(1-\lambda\,\alpha\,dt)\,P(0)+\frac{dt}{\bar{T}}P(1)=P(0),\\
&(\lambda\,\alpha\,dt)\,P(0)+\left[1-(\lambda-1)\,\alpha\,dt-\frac{dt}{\bar{T}}\right]P(1)\\
&\qquad+\frac{2\,dt}{\bar{T}}P(2)=P(1),\\
&\qquad\cdots\qquad\qquad\cdots\qquad\qquad\cdots\\
&(\lambda-\nu+1)\,\alpha\,dt\,P(\nu-1)+\left[1-(\lambda-\nu)\,\alpha\,dt-\frac{\nu\,dt}{\bar{T}}\right]P(\nu)\\
&\qquad+\frac{\nu\,dt}{\bar{T}}P(\nu+1)=P(\nu),\\
&(\lambda-\nu)\,\alpha\,dt\,P(\nu)+\left[1-(\lambda-\nu-1)\,\alpha\,dt-\frac{\nu\,dt}{\bar{T}}\right]P(\nu+1)\\
&\qquad+\frac{\nu\,dt}{\bar{T}}P(\nu+2)=P(\nu+1),\\
&\qquad\cdots\qquad\qquad\cdots\qquad\qquad\cdots\\
&\alpha\,dt\,P(\lambda-1)+\left[1-\frac{\nu\,dt}{\bar{T}}\right]P(\lambda)=P(\lambda),
\end{aligned}\right\}\qquad(201)$$

where $\alpha=\dfrac{n}{1-n\bar{T}-n\bar{\tau}}$.

These equations differ from those obtained in § 116 in just two respects: The first is, that sources are effectively busy,

so far as the origination of calls is concerned, whether they are actually being served, or are only awaiting service. Hence p, in (146), takes the form $n(\bar{T} + \bar{\tau})$, and gives for $n/(1 - p)$ the quantity which we have called α. But in the second elementary probability (148) p means, as before, $n\bar{T}$; for it is only when we find the source actually being served that there is a chance of that service coming to an end during dt.

In the second place, when the congestion exceeds the number of available channels, the number of calls in progress is equal to ν, not to j. Hence in place of (149) we must now write $j\,dt/\bar{T}$, provided $j \leqq \nu$, but $\nu\,dt/\bar{T}$ for all larger values of j.

From the first of the equations (201), $P(1)$ may be found in terms of $P(0)$. Then from the second $P(2)$ can be found in terms of $P(0)$, and so on.[1] As we should expect, the result takes different forms according as j is less than or greater than ν. It is

$$\left.\begin{aligned} P(j) &= C_j^\lambda\, \beta^j\, P(0), && j \leqq \nu \\ P(j) &= \frac{j!}{\nu!}\, \nu^{\nu-j}\, C_j^\lambda\, \beta^j\, P(0), && j \geqq \nu \end{aligned}\right\} \tag{202}$$

where β has been written for the quantity

$$\beta = \frac{n\bar{T}}{1 - n\,\bar{T} - n\bar{\tau}}. \tag{203}$$

So far the constant $P(0)$ in these equations is arbitrary, but it may be determined by means of the condition

$$\sum_0^\lambda P(j) = 1.$$

It is found to be given by the equation

$$\frac{1}{P(0)} = (1 + \beta)^\lambda + \sum_{j=\nu+1}^{\lambda} C_j^\lambda\, \beta^j \left(\frac{j!}{\nu!}\, \nu^{\nu-j} - 1\right). \tag{204}$$

These formulæ apply in case the number of sources is limited and the sources are independent. If, on the other

[1] The labor of solving (201) is much less if determinants are used than by the scheme suggested; but adeptness as well as knowledge is required.

hand, the calls occur individually and collectively at random, the solutions take somewhat simpler forms that may be obtained by setting λ infinite in the above expressions. They are

$$\left.\begin{aligned} P(j) &= \frac{\epsilon^j}{j!} P(0), \qquad j \leqq \nu \\ P(j) &= \frac{\epsilon^j}{\nu!} \nu^{\nu-j} P(0), \quad j \geqq \nu \end{aligned}\right\} \tag{205}$$

where ϵ is the expected traffic density, computed quite without regard to delays, and where $P(0)$ is found to be

$$\frac{1}{P(0)} = e^{\epsilon} + \frac{1}{(\nu - \epsilon)} \frac{\epsilon^{\nu}}{(\nu - 1)!} - \sum_{j=\nu}^{\infty} \frac{\epsilon^j}{j!}. \tag{206}$$

§ 139. *Exponential Distribution of Holding Times; The Expected Delay*

The probability $P(j)$ represents the proportion of time during which j calls are being served or are awaiting service. If j is not greater than ν, these calls are all being served, while if j exceeds ν, ν of them are obtaining service and the remaining $j - \nu$ are standing by waiting for an idle trunk. It follows, therefore, that the aggregate length of all calls served per unit time must be given by

$$\sum_{j=0}^{\nu} jP(j) + \sum_{j=\nu+1}^{\lambda} \nu P(j),$$

while the aggregate length of all the delays which occur per unit time must be

$$\sum_{j=\nu+1}^{\lambda} (j - \nu) P(j).$$

If the latter of these is divided by the expected number of calls per unit time, the expected delay, $\bar{\tau}$, is obtained. However, the average number of calls is $\lambda n = \epsilon/\bar{T}$. Hence we have the formula

$$\frac{\bar{\tau}}{\bar{T}} = \frac{1}{\epsilon} \sum_{j=\nu+1}^{\lambda} (j - \nu) P(j). \tag{207}$$

This can also be thrown into the form [1]

$$\frac{\bar{\tau}}{\bar{T}} = \frac{1}{\epsilon}\frac{\nu^{\nu}}{\nu!}\frac{P(0)}{P'(\lambda)^{*}}\left[\left(\lambda-\nu-\frac{\nu}{\beta}\right)(1-{}^{\infty}\Pi'^{*}_{\lambda-\nu})+\frac{\nu}{\beta}{}^{\infty}P'(\lambda-\nu+1)^{*}\right], \tag{208}$$

where the asterisk following a P or Π indicates that this symbol is to be evaluated as if the expected traffic density were ν/β.

From this equation it is quite possible to find $\bar{\tau}$, though it occurs implicitly on the right-hand side due to the relation (203). The process of computation consists in assigning to β such values as may be desired, computing the corresponding $\bar{\tau}$'s from (208), and then finding n from (203). In this way a table of corresponding values of n and $\bar{\tau}$ is obtained, from which the delay corresponding to any calling rate may be found by interpolation.

The computations are not at all simple; but they can be performed when it is necessary to answer questions of sufficient importance to justify the expense. In many cases it is satisfactory to assume that the calls originate individually and collectively at random, in which case the formula is much simpler. It is derived by exactly the same line of argument, and is found to be

$$\frac{\bar{\tau}}{\bar{T}} = \frac{1}{(\nu-\epsilon)^2}\frac{\epsilon^{\nu}}{(\nu-1)!}P(0). \tag{209}$$

§ 140. *Exponential Distribution of Holding Times; The Probability of a Delay Exceeding the Length τ if Calls are Served in the Order in which They Originate*

If a test call finds exactly j calls ahead of it, the delay to which it will be subjected will exceed the length τ if, and only if, $j-\nu$ or less calls end during the time τ. In case j is known, therefore, the probability of a delay as great as τ can be found

[1] In this case (as in several others which follow) we do not attempt to explain the method by means of which one form of equation follows from another. In each such case, however, the only processes required are those of routine algebra. They contain nothing of interest from a probability standpoint.

by determining the probability that a preassigned number of calls end during a given time τ. This is our next objective.

If we consider any time interval dt, at the beginning of which all trunks are busy, we know from § 138 that the probability that a call ends during this interval is $\nu\, dt/\bar{T}$. We also know from § 137 that the chance of a call ending during any such interval is altogether independent of what may have happened in any other. Hence these end-points are distributed individually and collectively at random with an expectation of $\nu/\bar{T}$ per unit length, and the chance that exactly i end during the time τ is

$$\left(\frac{\nu\tau}{\bar{T}}\right)^i \frac{e^{-\nu\tau/\bar{T}}}{i!}, \tag{210}$$

which we shall denote simply by ${}^{\infty}P'(i)^{**}$, the two asterisks indicating that the probability is to be computed as if the expected traffic density were $\nu\tau/\bar{T}$. The chance of $j-\nu$ or less terminating during this time (or, what amounts to the same thing, the probability of a delay exceeding τ) is

$$p_j(>\tau) = \sum_{i=0}^{j-\nu} {}^{\infty}P'(i)^{**} = 1 - {}^{\infty}\Pi'^{**}_{j-\nu+1}.$$

All this is true only provided we know that exactly j calls are either in progress or awaiting service when our test call occurs. However, during the proportion $P(j)$ of the time, j calls *are* in progress; and the calling rate during such intervals is $(\lambda - j)\alpha$ calls per unit time. It therefore follows that the number of calls originated per unit time which find just j calls preceding them and are delayed more than τ seconds is

$$(\lambda - j)\alpha\, P(j)\, p_j(>\tau).$$

This formula obviously applies for any value of j; and since any call which is delayed a longer time than τ must have been preceded by *some* number of calls, it follows that the number per unit time which may be expected to be delayed more than

τ may be obtained by summing this expression for such values of j as are capable of causing delays. The result is

$$\sum_{j=\nu}^{\lambda} (\lambda - j)\alpha\, P(j)\, p_j(> \tau).$$

When this result is divided by λn (that is, by the total number of calls which are expected to originate per unit time) it gives us the probability of a test call experiencing a delay greater than τ.

This formula can also be put in a somewhat more convenient form for purposes of computation. It is, in fact, equal to

$$P(>\tau) = \frac{\nu^{\nu}}{\epsilon(\nu-1)!} \frac{P(0)}{{}^{\infty}P'(\lambda)^*} \sum_{j=\nu}^{\lambda-1} {}^{\infty}P'(\lambda-j-1)^* \,(1 - {}^{\infty}\Pi'^{**}_{j-\nu+1}). \quad (211)$$

As in every other case, it becomes much simpler if we assume that the calls originate individually and collectively at random. The result is then

$$P(>\tau) = P(0)\, \frac{\epsilon^{\nu}}{(\nu - \epsilon)\,(\nu - 1)!}\, e^{(\epsilon-\nu)\tau/\overline{\tau}}. \quad (212)$$

It must be borne in mind that these results are only valid provided calls are served in the order in which they originate. If apparatus is assigned in some other fashion a different distribution of delays will occur. The results of §§ 139, 141 and 142, however, do not depend upon this assumption.

§ 141. *Exponential Distribution of Holding Times; The Proportion of Delayed Calls*

By setting τ equal to 0 in $P(>\tau)$ we obtain the proportion of delayed calls. From (211) we get

$$P(>0) = \frac{\nu^{\nu}}{\epsilon(\nu - 1)!}\, P(0)\, \frac{1 - {}^{\infty}\Pi'^{*}_{\lambda-\nu}}{{}^{\infty}P'(\lambda)^*}; \quad (213)$$

the asterisk having the same meaning as always. From (212) we get the even simpler formula

$$P(>0) = P(0)\, \frac{\epsilon^{\nu}}{(\nu - \epsilon)\,(\nu - 1)!}. \quad (214)$$

§ 142. *Exponential Distribution of Holding Times; The Expected Delay of Delayed Calls*

The expected delay, $\bar{\tau}$, obtained in § 139 is analogous to the result which would be obtained by placing a large number of test calls, noting the delay to which each was subjected, adding all these delays together, and dividing by the number of calls. Many of the calls, however, would be subject to no delay, and the average thus derived would apportion the total delay among these as well as among the delayed calls.

If we desire to know, not this expected delay $\bar{\tau}$ but the delay $\bar{\tau}_\delta$ which a call may be expected to have *if it is delayed at all*, we need only divide $\bar{\tau}$ by the proportion of delayed calls, $p(>0)$. The result is

$$\frac{\bar{\tau}_\delta}{\bar{T}} = \frac{1}{\nu} \frac{\left(\lambda - \nu - \frac{\nu}{\beta}\right)(1 - {}^{\infty}\Pi'^{*}_{\lambda-\nu}) + \frac{\nu}{\beta}\,{}^{\infty}P'(\lambda - \nu + 1)^{*}}{1 - {}^{\infty}\Pi'^{*}_{\lambda-\nu}}; \quad (215)$$

or, if calls occur individually and collectively at random,

$$\frac{\bar{\tau}_\delta}{\bar{T}} = \frac{1}{\nu - \epsilon}. \quad (216)$$

These results may complete our discussion. It is obvious that we might, by introducing various shades of meaning into our assumptions, prolong the study indefinitely, and that quite without the necessity of ever passing far from conditions for which we might easily find analogues in practice. What we have done, however, has probably shown the two things for which it was devised: the highly complex nature of the problems to which " traffic " in its most general sense leads us, and something — though by no means all — of the methods which in part at least meet the needs of such problems. If at the same time we have painted a rather wearisome picture, no great harm will have been done; for after all it was only the old masters to whom the gods granted a virgin field in

which to grow those things which were easiest, and to them no tools were given. We, who have inherited the implements of their fashioning, cannot well complain if the fields require more labor.

PROBLEMS

1. Prove (170) and (171).

2. What is the probability of hunting more than k terminals with a single selector starting from rest, and a straight multiple? What is the probability of hunting more than ν?

3. Prove that the exponential law (1) is the only distribution of holding times which possesses the property discussed in § 137.

(By keeping the expression for $p(> t)$ in terms of $p(t)$, an expression can be found for $p_{>t}(> t + dt)$ which is perfectly general; that is, it is true for any function $p(t)$. The property in question states that this quantity does not vary with t; which leads at once to a differential equation.)

4. Find the probability that a call is delayed, no matter how much, at a non-cooperative channel.

5. The $\epsilon(\tau)$ given by (198) is the "unconditional" expectation of delay. It is the analogue of an "average delay" formed from a number of calls, some of which suffered no delay at all. We could, however, eliminate from this group all those that suffered no delay, leaving only the "delayed calls"; and we could then make use of this residual group to find the "average delay of delayed calls." To such an average corresponds a "conditional expectation," namely, "the expectation of delay if the call is delayed," or in symbolical form $\epsilon_\delta(\tau)$. Find $\epsilon_\delta(\tau)$ at a non-cooperative channel.

References for Outside Reading

I. *General Studies:*

1. Erlang: *Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges*, Post Office Electrical Engineers' Journal, Vol. 10 (1918), pp. 189–197; *Telefon-Ventetider Et Stykke Sandsynlighedsregning*, Matematisk Tidsskrift (1920), pp, 25–42.

2. Engset: *Die Wahrscheinlichkeitsrechnung zur Bestimmung der Wählerzahl in automatischen Fernsprechämtern*, Elektrotechnische Zeitung (1918), pp. 304–305.

3. O'Dell: *Theoretical Principles of the Traffic Capacity of Automatic Switches*, Post Office Electrical Engineers' Journal, Vol. 13 (1920), pp. 209–223.

II. *Methods of Numerical Computation:*

4. Molina: *Computation Formula for the Probability of an Event Happening at Least C Times in N Trials*, American Mathematical Monthly, Vol. 20 (1913), pp. 190–193; *An Interpolation Formula for Poisson's Exponential Binomial Limit*, American Mathematical Monthly, Vol. 22 (1915), pp. 223–224.

5. Campbell: *Probability Curves Showing Poisson's Exponential Summation*, Bell System Technical Journal, Vol. 2 (1923), pp. 95–113.

CHAPTER XI

Fluctuation Phenomena in Physics

§ 143. *Introductory Remarks; Notation*

Among the many important applications of the Theory of Probability to scientific problems, those which deal with the Kinetic Theory of Gases and other statistical phenomena which arise from the molecular structure of matter and electricity form a class by themselves to which we may broadly apply the term "fluctuation phenomena." We have already had a few simple illustrations of such studies in §§ 64, 65, 66 and 88. It is the purpose of the present chapter to present a few other scattered results, with a view to illustrating the sort of thought processes that are required in this field.

To begin with, we shall give what is probably the most satisfactory derivation of Maxwell's Theorem that the velocities of gas molecules are distributed in accordance with the Normal Law. This derivation, like the one given in § 64, is founded upon a line of argument originally carried out by Maxwell himself; and while it is not satisfactory in all respects, as we shall see, it at least possesses the merit of leading to some very important physical and thermodynamic consequences. Before entering upon it directly, however, we have need to derive some dynamical information with which the reader may not be familiar.

As notation, we shall use the letters x, y, z to represent the coordinates of a point in space, and u, v, w for velocity components in the directions of the three coordinate axes. In order to avoid infinite repetition we shall often use the capitals X and U as substitutes for the triplets x, y, z and u, v, w, and shall write dX and dU in place of the more cumbersome

$dx\,dy\,dz$ and $du\,dv\,dw$. For example, we shall speak of a molecule as being in the element dX, meaning thereby that its center lies in the element of volume bounded by x and $x + dx$, y and $y + dy$, z and $z + dz$. Or we shall say that it is "of velocity class U," meaning that its velocity components lie within an infinitesimal region $du\,dv\,dw$ about the values u, v, w.

§ 144. *The Dynamics of Collision*

We consider two identical, perfectly hard, perfectly smooth and perfectly elastic spheres that move with velocities U and U'. We

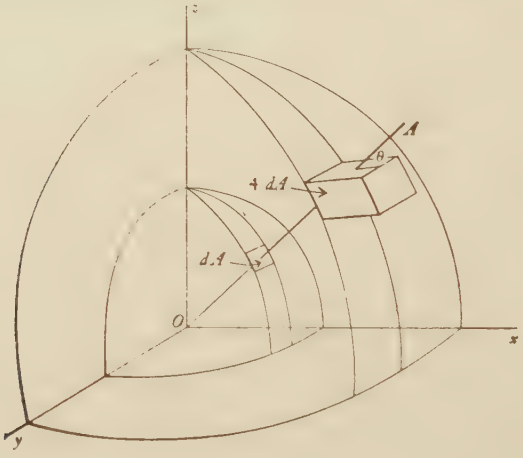


Fig. 48.

suppose that one of these is represented by the inner sphere of Fig. 48, and that the other collides with it, the point of contact being somewhere in the element of area dA. At the instant when the collision takes place, the center of the second sphere must be located on the surface of the outer sphere of Fig. 48; that is, the centers must be separated by a distance equal to the diameter of the spheres. Moreover, the center of the second sphere must be somewhere within the element of area $4\,dA$ which is homologous to the element dA upon the inner sphere. We ask, under these circumstances, for the velocities $\bar{u}$, $\bar{v}$, $\bar{w}$ and $\bar{u}'$, $\bar{v}'$, $\bar{w}'$ which these spheres will have after collision.

From the principle of the conservation of momentum we know

that what one sphere gains in momentum must be lost by the other. Hence we have

$$\left.\begin{aligned} \bar{u} - u &= u' - \bar{u}' = g_u, \\ \bar{v} - v &= v' - \bar{v}' = g_v, \\ \bar{w} - w &= w' - \bar{w}' = g_w, \end{aligned}\right\} \tag{217}$$

the g's being introduced merely for simplicity of expression.

Next, if the spheres are perfectly smooth, the acceleration must be in the direction of their line of centers OA at the moment of collision. We call the direction cosines of this line of centers λ, μ, ν. Hence we have

$$\frac{g_u}{\lambda} = \frac{g_v}{\mu} = \frac{g_w}{\nu} = S, \tag{218}$$

the S being again a constant introduced for simplicity.

Finally, energy must be conserved, so that

$$\bar{u}^2 + \bar{v}^2 + \bar{w}^2 + \bar{u}'^2 + \bar{v}'^2 + \bar{w}'^2 = u^2 + v^2 + w^2 + u'^2 + v'^2 + w'^2. \tag{219}$$

We now write the six equations (217) in the form

$$\begin{aligned} \bar{u} &= u + g_u, \\ \bar{u}' &= u' - g_u, \\ &\cdots \qquad \cdots \end{aligned}$$

square them, add the results together, and then subtract (219) from them. The result is

$$2(g_u^2 + g_v^2 + g_w^2) + 2[(u' - u)\, g_u + (v' - v)\, g_v + (w' - w)\, g_w],$$

or, by using (218) and remembering that the sum of the squares of the direction cosines is of necessity equal to unity,

$$2S^2 + 2S\,[\lambda(u' - u) + \mu(v' - v) + \nu(w' - w)] = 0.$$

It follows, then, that S must either be zero or take the value

$$S = \lambda(u - u') + \mu(v - v') + \nu(w - w'). \tag{220}$$

Whichever of these values S may have, the substitution of $g_u = \lambda S$, $g_v = \mu S$, $g_w = \nu S$ in (217) gives the desired relations

$$\left.\begin{aligned} \bar{u} &= u - \lambda S, & \bar{u}' &= u' + \lambda S, \\ \bar{v} &= v - \mu S, & \bar{v}' &= v' + \mu S, \\ \bar{w} &= w - \nu S, & \bar{w}' &= w' + \nu S. \end{aligned}\right\} \tag{221}$$

It is now obvious that the value $S = 0$ is not the true one, for it would require the velocities after contact to be the same as before.

We have now derived the equations which define the new velocities in terms of the old ones and of the direction cosines of the line of centers at the instant of collision. But we still want a geometrical meaning for the letter S. This we get by noting that the velocity of the sphere of class U relative to the one of class U' has the velocity components $u - u'$, $v - v'$, $w - w'$. Its absolute value is therefore $R = \sqrt{(u - u')^2 + (v - v')^2 + (w - w')^2}$ and its direction cosines are the ratios of the three components to this absolute value R. If we denote these direction cosines by λ_R, μ_R, ν_R, and substitute them in (220) we find that

$$S = [\lambda\lambda_R + \mu\mu_R + \nu\nu_R]R.$$

However, by an elementary theorem in analytic geometry, the combination of direction cosines which occurs in the brackets is equal to the cosine of the angle included between the directions in question; that is, between the line of centers at the moment of collision, and the direction of the relative velocity. If we call this angle θ, (220) takes the simpler form

$$S = R \cos \theta. \tag{222}$$

In other words, the symbol S represents the projection of the relative velocity upon the line of centers.

We shall next show that, if we were to start with two spheres having velocities $\bar{U}$ and U' (that is, the velocities with which the other pair *emerged* from their collision), and if we allowed them to collide in such a way that, at the instant of collision, the line of centers was the same as before, they would emerge with velocities U and U' (that is, with velocities equal to those which the other pair had *before* collision).

Of course, since the argument by means of which we obtained (221) is a perfectly general one, it follows that the velocity components of our new set of spheres, after collision, must be given by the equations

$$\bar{\bar{u}} = \bar{u} - \lambda\bar{S}, \quad \bar{\bar{u}}' = \bar{u}' + \lambda\bar{S}, \quad \ldots, \tag{223}$$

in which symbols such as $\bar{\bar{u}}$ represent the desired velocities after collision, and $\bar{S}$ bears the same relation to our new velocities that S did to the old ones. As the line of centers is the same as before, the direction cosines remain unchanged.

The quantity S is the central element in our argument. By

merely changing the symbols in (220) so as to apply to our present case we find that it must satisfy the relation

$$\bar{S} = \lambda(\bar{u} - \bar{u}') + \mu(\bar{v} - \bar{v}') + \nu(\bar{w} - \bar{w}').$$

But from (221) we have $\bar{u} - \bar{u}' = u - u' - 2\lambda S$, with entirely similar equations in the v's and w's. Making use of these we easily throw the equation for $\bar{S}$ into the form

$$\bar{S} = \lambda(u - u') + \mu(v - v') + \nu(w - w') - 2(\lambda^2 + \mu^2 + \nu^2)S,$$

which is obviously equivalent to $\bar{S} = -S$.

If we now substitute this value of $\bar{S}$ in (223) and compare the resulting equations with (221), we find that U and U' are indeed identical with $\bar{U}$ and $\bar{U}'$, which is what we set out to prove.

§ 145. *The Probable Flux Across a Surface*

Let us now consider a set of spheres moving about in the fashion in which the gas molecules are supposed to move in the Kinetic Theory. We shall suppose these spheres, or molecules, to be hard and smooth and elastic, just as we did in § 144, and we shall further agree to represent by the formula $p(u, v, w, x, y, z, t)\, du\, dv\, dw\, dx\, dy\, dz$ the chance of a molecule lying within the parallelopiped bounded by the planes $x, x + dx$; $y, y + dy$; $z, z + dz$ and having velocity components that lie between u and $u + du$, v and $v + dv$, $w + dw$ respectively. It is, of course, our purpose to find what this function is; but our way of arriving at the answer will be a somewhat indirect one, and for the moment the symbol itself is quite sufficient. It should be noted, however, that we are not assuming that one element of volume is just as likely to contain a molecule as another, for we have made p a function of x, y, z; nor are we assuming that the gas is in a state of statistical equilibrium, for we have made p a function of time.

We start, then, with this symbol $p(U, X, t)\, dU\, dX$ for the chance of there being a molecule of class U in the element dX at the time t, and ask for the probability of such a molecule passing out between the time t and $t + dt$, dt being supposed infinitesimally small. We shall represent this probability by $P(\overline{X})\, dU\, dX\, dt$.

We refer to Fig. 49 and note that, if a sphere crosses the boundary of dX at all, it must cross one of the six faces. We shall therefore find the probability we desire if we consider each of these six faces separately. Let us take first the pair which are perpendicular to the x-direction.

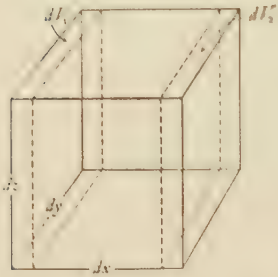


FIG. 49.

A molecule of class U cannot leave this volume element across the left-hand face if u is positive; if u is negative it cannot leave across the right-hand face. Hence the chance of the molecule leaving across one of these faces in time dt is just the chance of it lying in some such slab as dV_2 if u is positive, or dV_1 if u is negative. We thus find that the chance of a molecule leaving during dt is just

$$p(u, v, w, x + dx, y, z, t)\, du\, dv\, dw(u\, dt)\, dy\, dz$$

if u is positive, and

$$p(u, v, w, x, y, z, t)\, du\, dv\, dw\, (-u\, dt)\, dy\, dz$$

if u is negative, the use of $x + dx$ in the former of these expressions being dictated by the fact that the slab lies adjacent to the right-hand face of dX.

Of course there are entirely similar expressions for the chances of leaving across either of the other pairs of faces, and by summing these expressions we might find $P(\overline{X})$. The result would be a rather complicated one, however, as it would have different forms for each of the eight possible combinations of signs of u, v, w. Fortunately, we have no use for the formula in just this form, and need not carry the matter into greater detail. Instead, we shall turn our attention to the chance of *entering* dX in this same way, which we shall call $P(X)\, dU\, dX\, dt$.

The argument for this case is just the same as was the argument for leaving. The only difference is that the slab within which the molecule must lie in order to cross within dt

is now *outside* dX instead of inside, and is adjacent to the opposite face in every instance. It follows that the chance of entering across the pair of faces which we have been considering is

$$p(u, v, w, x, y, z, t)\, du, dv, dw\, (u\, dt)\, dy\, dz$$

if u is positive, and

$$p(u, v, w, x + dx, y, z, t)\, du\, dv\, dw\, (-u\, dt)\, dy\, dz$$

if u is negative.

We shall not be much interested in either $P(X)$ or $P(\overline{X})$ directly, but we shall need to know the difference $P(X) - P(\overline{X})$, which may readily be found by discussing one pair of faces at a time and adding the results. For the pair which we have had under consideration, we readily see that the difference is $[p(u, v, w, x, y, z, t) - p(u, v, w, x+dx, y, z, t)]u\, du\, dv\, dw\, dy\, dz\, dt$, no matter whether u be positive or negative. As dx itself is supposed to be infinitesimally small, this may be rewritten in the form

$$-\frac{\partial p}{\partial x}\, u\, dU\, dX\, dt.$$

It is obvious that the analogous probabilities for the other pairs of faces are $-\frac{\partial p}{\partial y}\, v\, dU\, dX\, dt$ and $-\frac{\partial p}{\partial z}\, w\, dU\, dX\, dt$, whence we obtain as our final result

$$P(X) - P(\overline{X}) = -\left(u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y} + w\frac{\partial p}{\partial z}\right), \qquad (224)$$

the differential elements being $dU\, dX\, dt$ on both sides of the equation.

§ 146. *Change of Velocity Class*

Another thing in which we shall be interested in later sections is the chance of a molecule changing from the velocity class U to some other velocity class during the time dt, and the analogous chance of it changing from some other velocity class to the class U.

The chance of a sphere of class U being within the element of volume dX at the time t is $p(U, X, t)\, dU\, dX\, dt$. If a sphere of class U' comes along and collides with it during the time interval dt, this latter sphere must have been located somewhere within a well-defined volume element when the time interval dt began. Specifically, it must have been within that element which the area $4\, dA$ (Fig. 48) would sweep out, if it were to travel for the time dt with a velocity equal to the velocity R of the U-class sphere relative to the U'-class sphere.

It is not difficult to see that the volume of this element is just $4\, R \cos\theta\, dA\, dt$, where θ is the angle between the relative velocity R and the line of centers at the instant of collision. From (222) we see at once that it may be written in the alternative form $4\, S\, dA\, dt$.

Now the chance of a collision of the sort under consideration is just the product of the probability of a U-class sphere in dX, which we know, by the *conditional* probability that there is a U'-class sphere in the other volume element *if there is a U-class sphere in dX*. What this conditional probability is we do not know. Moreover, no way has ever been found to determine it without advance knowledge of the distribution function $p(U, X, t)$, which we obviously do not possess.[1] Hence we seem to be blocked from further exact progress. If, however, we *assume* that the existence of a U'-class molecule in any element of volume is not in any way affected by the proximity of U-class molecules — in other words, if we assume the two events to be *independent* — we may make use of the *uncon-*

[1] Jeans, in his *Dynamical Theory of Gases*, gives a development of Maxwell's Law which he believes to be free from this objection. The demonstration is phrased in terms of Statistical Mechanics, and while it does not need our assumption in the exact form in which we have made it, it appears to me that something very similar lurks in assuming a uniform density for the "dust of points" in his statistical space of $6N$ dimensions.

On the other hand, Jeans has shown that if the gas is very tenuous and is distributed in accordance with Maxwell's Law, the chance of a U'-class molecule in any element is independent of the situation in neighboring elements; which seems to complete the argument in favor of Maxwell's Law *a posteriori* for such rare gases. For gases in which the molecules occupy an appreciable portion of the available space he has shown that the condition of independence does not exist.

ditional probability of a U'-class molecule in the element $4\,S\,dA\,dt$ and arrive at a result; for that unconditional probability is just $p(U', X', t)\,dU'\,4\,S\,dA\,dt$, where X' denotes some point located in the element $4\,S\,dA\,dt$. Therefore the chance of a collision occurring between two such molecules is

$$4\,S\,p(U, X, t)\,p(U', X', t)\,dU\,dX\,dU'\,dA\,dt. \qquad (225)$$

This is the chance of a collision of a special kind. We must not forget, however, that we are seeking for the chance of a molecule leaving the class U, and this result will follow from any collision whatever.[1] Hence it is necessary for us to sum the probability (225) over every class of molecule with which our U-class molecule could possibly collide. The result is

$$P(\bar{U})\,dU\,dX\,dt = 4\,dU\,dX\,dt \int dU' \int dA\,S\,p(U, X, t)\,p(U', X', t). \qquad (226)$$

This is the chance of a molecule *leaving* class U by collision. As for the chance of a molecule *entering* class U, that can best be obtained by indirection. We saw in § 144 that if the colliding molecules had velocities $\bar{U}$ and $\bar{U}'$, and collided so that their line of centers had the appropriate direction, they would give rise to two new velocities U and U'. To get the chance of a molecule entering class U in just this way, therefore, we need only multiply together the probabilities of their having had, before collision, velocities which lay within the ranges to which collisions between U- and U'-class molecules might lead. Let us denote these ranges by $d\bar{U}$ and $d\bar{U}'$. Then we have at once, for the probability of a collision of exactly the type about which we have been speaking,

$$4\bar{S}\;p(\bar{U}, X, t)\,p(\bar{U}', \bar{X}', t)\,d\bar{U}\,dX\,d\bar{U}'\,dA\,dt;$$

[1] So long as the differential element dU is of finite magnitude it is possible for the molecule to undergo a collision which changes its velocity by so little that it still remains in the same class. The chance of this vanishes with dU, however, and does not invalidate our argument.

and therefore, for the total probability of entering class U through any sort of collision whatever,

$$P(U)\,dU\,dX\,dt = 4\,dU\,dX\,dt \int dU' \int dA\,S\,p(\bar{U}, X, t)\,p(\bar{U}', \bar{X}', t). \tag{227}$$

We have seen in § 144, however, that $S = -\bar{S}$; and we have seen in § 68 that the Jacobian of the transformation (221) is equal to unity, which means that $dU\,dU' = d\bar{U}\,d\bar{U}'$. We conclude at once, therefore, that (227) may be written in the form[1]

$$P(U)\,dU\,dX\,dt = 4\,dU\,dX\,dt \int dU' \int dA\,S\,p(\bar{U}, X, t)\,p(\bar{U}', \bar{X}', t). \tag{228}$$

Again, as in § 145, we shall be interested less in $P(U)$ and $P(\bar{U})$ separately than in their difference. We shall therefore subtract (228) from (226) and write

$$P(U) - P(\bar{U}) = 4 \int dU' \int dA\,S\,[\bar{p}\bar{p}' - pp']. \tag{229}$$

Strictly speaking, the symbols p and $\bar{p}$ in (229) refer to the chance of a molecule being within an element of volume located at the point X, while p' and $\bar{p}'$ refer to the probabilities at points situated a distance from X equal to the diameter of a molecule. This is obvious from Fig. 48. But the molecules are so small that $p(U', X', t)$ and $p(\bar{U}', \bar{X}', t)$ will generally differ by negligible amounts from $p(U', X, t)$ and $p(\bar{U}', X, t)$. In what follows, therefore, we shall assume that all the symbols in (229) refer to the point X.

[1] We have dropped the sign of S for just the same reason as we ignore the sign of the Jacobian. It is the result of certain conventions as to directions and is more easily corrected in this arbitrary fashion than kept correct. In the present instance we know that all quantities must of necessity be positive.

§ 147. *The Fundamental Equation of the Kinetic Theory of Gases*

We are now prepared to derive the integro-differential equation which forms the mathematical basis for the entire Kinetic Theory of Gases. We suppose that an element of volume dX, is taken under observation at the instant t, and that the observation is extended to the time $t + dt$. We ask for the probability that, at the *end* of this interval, the element contains a molecule moving with the velocity U. In other words, we ask for the probability $p(U, X, t + dt)$.

This probability, however, can be expressed as the sum of five [1] related probabilities all of which have already been found. They are:

1. The probability that a molecule of velocity class U was in dX at the time t.[2] This is, of course, $p(U, X, t)$.

2. The chance that such a molecule was in dX at time t but wandered across the boundary during dt. We have already denoted this by the symbol $P(\overline{X})\, dU\, dX\, dt$.

3. The chance that such a molecule was in dX at time t, but suffered a collision which caused it to alter its velocity. We have denoted this by $P(\overline{U})\, dU\, dX\, dt$.

---

[1] If the molecules are subject to extraneous forces, such as gravitational forces, for example, it is possible for a molecule to enter or leave class U by acceleration. If we were to take account of these possibilities we would have seven items, instead of five, to consider. The result would be to add to the left-hand side of (230) three additional terms

$$F_x \frac{\partial p}{\partial u} + F_y \frac{\partial p}{\partial v} + F_z \frac{\partial p}{\partial w},$$

F_x, F_y and F_z being the components of the applied force in the three coordinate directions. This would lead to a certain amount of algebraic complication and would cause us to reach a different distribution function p: the gas in a large tank is denser at the bottom and rarer at the top than it would be if the earth exerted no gravitational attraction, for example. But it would lead to no new ideas of a statistical nature and is not in place here. We assume, therefore, that our gas is not under the influence of such extraneous forces.

[2] To be strictly accurate we should add "and remained there." But if dt is infinitesimal, the chance of it *not* remaining there is also infinitesimal, and may be ignored.

4. The chance that a molecule of velocity class U was near enough dX at the time t to wander *into* it during dt. We have denoted this by $P(X)\,dU\,dX\,dt$.

5. The chance that a molecule of velocity class other than U was in dX at time t, and suffered a collision during dt which caused it to enter the class U. We have denoted this probability by $P(U)\,dU\,dX\,dt$.

Adding together these five probabilities, as given by (229) and (224), and observing that

$$p(U, X, t + dt) - p(U, X, t) = \frac{\partial p}{\partial t}\,dt,$$

we arrive at the formula

$$\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y} + w\frac{\partial p}{\partial z} = 4\int dU' \int dA\, S\,[\bar{p}\bar{p}' - pp']. \quad (230)$$

This is the fundamental equation of the Kinetic Theory. It is an integro-differential equation, among the solutions of which must be found the distribution function $p(U, X, t)$ for which we are seeking; that is, the distribution of velocities in a gas which is in statistical equilibrium. But it is much more general than the result which we shall obtain from it; for as we noted at the beginning; we have throughout taken account of the possibility of time variation. If then, we were to put a gas in some state other than equilibrium and then leave it to itself, the distribution function which governed its return to equilibrium would have to be given, instant by instant, by some one of the solutions of (230). From this equation, then, we should be able to learn many things about such readjustments within a gas — the time required to carry them out, for instance; and it has actually proved very useful in the treatment of such problems, which are, however, too technical for an elementary text in the Theory of Probability.

§ 148. *The H-Function*

So far, we have spoken only of the chance of an element of volume possessing a molecule of a certain kind. We now desire, for a time, to talk of the chance of a molecule being in

a certain state at a certain time. For this purpose, let us suppose that there are within the container N molecules and that we have some means of identifying them from one another. Further, we assume that, if there is a molecule of class U in dX at time t, it is just as likely to be any one of the set as any other.

We denote by $p^*(U, X, t)$ the chance of our particular molecule being in such a place and state at time t. Then obviously

$$p^*(U, X, t) = \frac{1}{N} p(U, X, t). \tag{231}$$

It is, of course, the same for every molecule.

Let us now compute the expectation of the logarithm of this probability, after the fashion explained in § 80. Formally, at least, the result is

$$\epsilon_1(\log p^*) = \int dU \int dX\, p^*(U, X, t) \log p^*(U, X, t).$$

Using (231), and remembering that the integral of $p^*(U, X, t)$ over all possible places and velocities must of necessity equal unity, we can easily reduce this to the form [1]

$$H(t) = N\, \epsilon_1(\log p^*) = -N \log N + \int dU \int dX\, p \log p. \tag{232}$$

Had we chosen a time $t + dt$ instead of a time t we would have gotten an entirely analogous result, except that in every instance the value of t would have been different. It follows, then, that $H(t)$ must satisfy the differential equation

$$\frac{dH}{dt} = \int dU \int dX\, \frac{\partial p}{\partial t} (\log p + 1), \tag{233}$$

which is obtained by straightforward differentiation of (232).

[1] The exact way in which the H is defined, and indeed the use of the expectation of the *logarithm* of the probability at all, is dictated by classical usage in the Kinetic Theory of Gases. The interpretation which we give to it, however, is not the usual one. The latter will be found in any good treatise on the Kinetic Theory.

We have already obtained an expression for $\partial p/\partial t$ in § 147, and upon substituting this in (233) we have

$$\frac{dH}{dt} = +\int dU \int dX \int dU' \int dA\, S\,(\bar{p}\bar{p}' - pp')(\log p + 1)$$

$$-\int dU \int dX \left(u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y} + w\frac{\partial p}{\partial z}\right)(\log p + 1). \quad (234)$$

The last term of this equation is readily thrown into the form

$$-\int dU \int dX \left(u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}\right)(p \log p),$$

which is known by Green's Theorem to be just the surface integral, over the entire container of the gas and for all possible values of the velocity, of the product of the *normal* component of U by the quantity $p \log p$. The product $p \log p$, however, has, *at the surface of the container*, the same value for a negative normal component as for a positive one; for if the walls of the container are not in motion every molecule which strikes leaves with its normal velocity reversed. Hence the entire expression is an odd function of this normal component, and in summing over every possible value of U it occurs as often with a positive as with a negative sign. The entire integral is therefore zero.

We conclude, then, that dH/dt is represented by the first term of (234) only.

Finally, we notice that, had we chosen from the very beginning to speak of our molecule as belonging to class U' instead of U, we would have obtained an entirely similar expression, the only difference occurring in the substitution of $\log p' + 1$ in place of $\log p + 1$. If we had chosen to speak of a molecule of class $\bar{U}$ we would have gotten

$$\frac{dH}{dt} = +\int d\bar{U} \int dX \int d\bar{U}' \int dA\, \bar{S}\,(pp' - \bar{p}\bar{p}')(\log \bar{p} + 1).$$

But since $\bar{S}\, d\bar{U}\, d\bar{U}' = S\, dU\, dU'$, as we saw in § 146, this becomes

$$\frac{dH}{dt} = -4\int dU \int dX \int dU' \int dA\, S\,(\bar{p}\bar{p}' - pp')(\log \bar{p} + 1).$$

There is also a similar expression which might have been derived had we chosen to speak of a molecule being of class $\bar{U}'$. It differs only in the replacement of $\log \bar{p} + 1$ by $\log \bar{p}' + 1$. Adding the four expressions for dH/dt thus obtained we reach the more symmetrical result

$$\frac{dH}{dt} = \int dU \int dX \int dU' \int dA\, S\,(\bar{p}\bar{p}' - pp')(\log pp' - \log \bar{p}\bar{p}'). \tag{235}$$

We may now notice this fact: When $\bar{p}\bar{p}'$ is greater than pp', the logarithm of the former exceeds the logarithm of the latter, and conversely. The two factors in the integrand of (235) are therefore always opposite in sign (unless they are zero), whence, since S is positive by definition, the entire integral must be negative. The only possible exception occurs when the expression $\bar{p}\bar{p}' - pp'$ is *identically* zero.

We conclude that $\dfrac{dH}{dt}$, and therefore $\dfrac{d}{dt}\epsilon_1(\log p^*)$ also, is never positive. In other words, *in any dynamical system of "molecules" of the sort under discussion, the expectation of* $\log p^*(U, X, t)$ *is a decreasing, or at least not an increasing, function of the time.*

Now we have seen in § 80 that the expectation of $\log p$ is least when the distribution of the variable is completely random. The fact that the $\epsilon_1(\log p^*)$ is continually decreasing therefore indicates a tendency of the gas, when left to itself, to approach more and more nearly to a condition of absolute randomness. We might, indeed, define H as measuring the "departure from randomness." Then when H reached its minimum value the gas would be as near to a completely random distribution as the dynamical conditions by which it is surrounded will permit.

§ 149. *Maxwell's Law of Velocities*

Let us now find what distribution of positions and velocities is as nearly random as the dynamical conditions of the problem will permit. The only essential dynamical conditions of which we know are, that the number of molecules must always be the same, and that the total energy shall not vary. Certainly if the number of molecules does not change, the expectation of that number must have the fixed value N, whence we get

$$N = \int dU \int dX\, p. \tag{236}$$

And if the total energy is a constant, the expectation of the total energy must be equal to the same constant. It is usual in physics to denote this constant by $\frac{3}{2}kNT$, N being the number of molecules, T the temperature and k a constant determined by the mechanical equivalent of heat. If we use this same notation, our function p must satisfy the relation

$$W = \tfrac{3}{2}\, kNT = \frac{m}{2} \int dU \int dX\, (u^2 + v^2 + w^2)\, p, \tag{237}$$

m being the mass of the molecule.

We are now required to make

$$H = \int dU \int dX\, p \log p$$

a minimum subject to the conditions that N and W shall remain unchanged. Following out the process explained in § 80, we find that it is necessary to make the integral

$$\int dU \int dX\, p\, [\log p - \lambda - \mu\, (u^2 + v^2 + w^2)]$$

as small as possible, without restrictions. If, however, we replace p by $p + \delta$ and require that the integral which contains

the first power of δ shall vanish identically, we get the solution

$$\log p = \lambda - 1 + \mu(u^2 + v^2 + w^2), \tag{238}$$

or

$$p = C\, e^{\mu(u^2+v^2+w^2)}. \tag{239}$$

There remains the matter of determining the arbitrary constants C and μ. For this purpose we have the two relations (236) and (237), both of which must be satisfied by our function. Substituting (239) in (236), we observe at once that the integrations with respect to u, v, w will give infinite results unless μ is negative. But if μ is negative each of the three leads to a factor $\sqrt{-\pi/\mu}$. This causes (236) to take the form

$$N = C\left(-\frac{\pi}{\mu}\right)^{3/2} \int dX.$$

As the integral of dX over all possible values is just the volume of the container, we get

$$\frac{N}{V} = C\left(-\frac{\pi}{\mu}\right)^{3/2}$$

as our first condition.

By substituting (239) in (237) and carrying out an entirely similar set of integrations we arrive at another relation

$$\frac{N}{V}\, kT = \frac{m}{2}\, C\, \frac{\pi^{3/2}}{(-\mu)^{5/2}}.$$

If we solve these equations and set the ratio N/V, which is the number of molecules per unit volume, equal to ν, we get the results

$$\left.\begin{aligned} \mu &= -\frac{m}{2kT} = -a, \\ C &= \nu\left(\frac{m}{2\pi\, kT}\right)^{3/2} = \nu\left(\frac{a}{\pi}\right)^{3/2}. \end{aligned}\right\} \tag{240}$$

Let us now assure ourselves that this solution really satisfies (230), for we must remember that we have obtained it, not by

solving (230) itself, but by making H a minimum. By substituting (238) in (219) we obtain

$$\log p + \log p' = \log \bar{p} + \log \bar{p}',$$

or what amounts to the same thing

$$pp' = \bar{p}\bar{p}'.$$

Hence the right-hand side of (230) vanishes. That the left-hand side also vanishes is obvious at once, since p is a function of neither x, y, z nor t.

Finally, let us compare our solution with (73). We remember, to begin with, that (73) is a formula for the probability of a molecule being in a certain state, as is obvious from the use of the asterisk. Moreover, it represents the chance of the molecule having the velocity U, *no matter where it may be*, whereas the $p^*(U, X, t)$ of (231) is the probability of a molecule being in a particular place and having a particular velocity. The relation between the two is obviously

$$p^*(u, v, w) = \int p^*(U, X, t)\, dX = \frac{V}{N} p(U, X, t).$$

Hence we get, from (239) and (240),

$$p^*(u, v, w) = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-\frac{m}{2kT}(u^2+v^2+w^2)}, \qquad (241)$$

which is indeed identical with (73) provided we make $a = m/2kT$, as we have done in writing (240).

§ 150. *Pressure*

When a molecule collides with the wall of the containing vessel, its component of momentum is reversed. To do this requires the exertion of a force upon the molecule, and of course imposes an equal force upon the wall. Theoretically, since we are assuming that both the molecules and the walls are perfectly hard, the change of momentum takes place instantaneously, and therefore the force necessary to bring it about must be infinite. In other words, at those instants

when molecules collide with the walls of the vessel, it is subjected to an infinite force; at all other instants it is subject to no force whatever.

We do not ordinarily think of the pressure of a gas as being of this nature: we think of a toy balloon, for instance, as being stretched by the steady exertion of an unvarying force upon every element of its surface. But in the Kinetic Theory the pressure is supposed to be the result of all these collisions, and to appear steady only because we are incapable of observing the individual impulses. In other words, in the Kinetic Theory "pressure" is the expectation of the force per unit area of the surface.

We may arrive at our result as follows: If a variable force acts for a time T, its average value is, by definition

$$\bar{f} = \frac{1}{T}\int_0^T f\,dt.$$

Suppose, now, that such a force acts in the direction of the x-axis upon a body of mass m. Then the relation between force and acceleration shows us that

$$f = m\frac{du}{dt},$$

from which we readily find that

$$\bar{f}\,T = m\,u_T - m\,u_0,$$

u_0 and u_T being the velocities of the particle at the beginning and end of the interval.

Stated in words this equation reads, "the average of a variable force is equal to the change of momentum produced in time T, divided by T." The exact analogue of this, expressed in terms of expectation instead of average is:

"The pressure exerted upon the walls of a vessel containing a gas is equal to the expectation of the change of momentum per unit area per unit time."

As the pressure is the same in all directions, we may as well consider an element of area perpendicular to the axis of x.

The probability of a molecule striking such an element with a velocity U in time dt is obviously

$$p(U, X, t)\, u\, dA\, dt,$$

dA being the area of the element, and u being allowed to have only positive values since there is gas on only one side of the wall. Each such molecule undergoes a change of momentum equal to $2mu$. Hence the product of this change of momentum by its probability of occurrence is $2mu^2\, p(U, X, t)\, dA\, dt$; and summing this over every possible value of U (that is, over every value for which u is positive), we get

$$P\, dA\, dt = 2m\nu\left(\frac{m}{2\pi kT}\right)^{3/2} dA\, dt \int_0^\infty du \int_{-\infty}^{\infty} dv \int_{-\infty}^{\infty} dw\, u^2\, e^{-\frac{m}{2kT}(u^2+v^2+w^2)}$$

which works out to be $P = \nu kT$.

If we replace ν by N/V this becomes the well-known law of perfect gases

$$PV = kNT:$$

the product of pressure by volume is proportional to temperature.

§ 151. *The Expectation of the Distance Travelled by a Molecule*

From the probability (76) of a molecule having a given speed we can readily find the expectation of the distance travelled in time dt. For if the molecule is travelling with speed s it will travel a distance $s\, dt$ during such an interval. The expectation of this distance is therefore

$$4\, dt\, \frac{a^3}{\pi} \int_0^\infty s^3\, e^{-as^2}\, ds = \frac{2}{\sqrt{\pi a}}\, dt.$$

As the expectation of the distance is proportional to the time, no matter how long or how short the interval may be, it is not necessary to regard dt as an infinitesimal. We may therefore say: In unit time a molecule may be expected to travel a distance

$$D = \frac{2}{\sqrt{\pi a}} = 2\sqrt{\frac{2kT}{\pi m}}. \tag{242}$$

§ 152. *Number of Collisions*

We have found in § 146 that the chance of a collision between a U-class molecule and one of some other kind within the element dX and during the time dt, is given by (226). The expectation of the total number of collisions in the entire container during dt is therefore the integral of this expression over all possible values of U and X. But if we compute the integral in this way we really count each collision twice, for if we think of two special values U_1 and U_2 we count not only $U = U_1$ and $U' = U_2$, but also $U = U_2$ and $U' = U_1$. We must therefore take *half* the integral in order to get the correct expectation. That is

$$\epsilon = 2\,dt \int dU \int dX \int dU' \int dA\, S\, p(U, X, t)\, p'(U', X', t).$$

The evaluation of this integral is a rather complicated piece of calculus, and it will be sufficient for us to state only the final result, which is

$$8\,\nu^2\,\rho^2\,V\sqrt{\frac{\pi kT}{m}} \tag{243}$$

where ρ is the radius of a molecule.

This, of course, is the total number of collisions that may be expected to take place in the entire volume V. To find the number in which any one molecule may be expected to partake, we may notice that if each collision involved only *one* molecule this would be just $1/N$ times (243). But since each collision involves *two* molecules, the true answer is twice as great as this. That is, each molecule can be expected to take part in

$$16\,\nu\,\rho^2\sqrt{\frac{\pi kT}{m}} \tag{244}$$

collisions per unit time.

§ 153. *Expected ("Mean") Free Path*

We next ask how far a molecule may be expected to travel between collisions. Having already found the distance it may

be expected to travel per unit time, and the number of collisions it may be expected to have within that distance, the expected free path can be very easily obtained by dividing (242) by (244). The result is

$$\frac{\sqrt{2}}{4\pi\nu\rho^2}.$$

This is as far as we are justified in discussing the Kinetic Theory in an elementary text, and indeed we have introduced the major part of the mathematical ideas which underlie it. Beyond this point it becomes largely an application of these ideas to the discussion of various physical systems. We leave the Kinetic Theory, therefore, for a few simple problems of a slightly different type.

§ 154. *Density Fluctuations*

Problems of a sort related to those of the Kinetic Theory appear in various branches of science. One of the simpler ones is concerned with what may be briefly characterized as "density fluctuations." It appears in many places. For example:

Gas molecules are continually wandering into and out of any element of volume which we may choose to take under observation. The number in it therefore varies with the time, or, in other words, the density of the gas fluctuates from point to point and from time to time.

If we blow some fine dust particles into a gas-filled vessel, they will wander about in a haphazard fashion because of the impacts which they receive from the molecules of the gas. The number of dust particles within our element of observation will also vary with the time. That is, the dust "density" will vary.

If we observe a thin film of liquid under the microscope we may see colloidal particles or living organisms wandering about in it. If so, the number within any marked portion of the field of vision fluctuates with the time.

Exactly this same phenomenon appears in so many ways

that it is natural to ask as to the nature of the fluctuations. Specifically, therefore, we shall seek to find the probability that a random observation will show exactly j particles within the observed element of volume.

To solve this problem, let us introduce the symbol $p_j\,dt$ to represent the chance of a particle entering the element during the interval dt, and $\bar{p}_j\,dt$ to represent the probability of one leaving, the subscript j indicating the presence of j particles at the beginning of the interval. Then if $P(j, t)$ represents the probability of just j particles at the instant t, it is a simple matter to set up the equations

$$P(0, t + dt) = P(0, t)(1 - p_0\,dt) + P(1, t)\,\bar{p}_1\,dt,$$

$$P(j, t + dt) = P(j - 1, t)\,p_{j-1}\,dt + P(j, t)(1 - p_j\,dt - \bar{p}_j\,dt) + P(j + 1, t)\bar{p}_{j+1}\,dt,$$

$$P(\lambda, t + dt) = P(\lambda - 1, t)\,p_{\lambda-1}\,dt + P(\lambda, t)(1 - \bar{p}_\lambda\,dt);$$

the last of these being written on the assumption that only λ particles are available, so that the chance of a larger number in the element is zero. If the number may be regarded as infinite, the last equation need not be considered.

We can readily reduce these to the form of *differential* equations and obtain

$$\left.\begin{aligned} \frac{dP(0)}{dt} &= -\,p_0\,P(0) + \bar{p}_1\,P(1). \\ \frac{dP(j)}{dt} &= p_{j-1}\,P(j-1) - (p_j + \bar{p}_j)\,P(j) + \bar{p}_{j+1}\,P(j+1), \\ \frac{dP(\lambda)}{dt} &= p_{\lambda-1}\,P(\lambda - 1) - \bar{p}_\lambda\,P(\lambda). \end{aligned}\right\} \quad (245)$$

It is no longer necessary to write the time explicitly, since we shall have no further occasion to refer to the instant $t + dt$.

If we now assume that the system is in statistical equilib-

rium, the derivatives vanish from (245), leaving a set of linear forms the solution of which can readily be found. It is,

$$P(j) = \frac{\Delta_j}{\sum \Delta_j}, \tag{246}$$

where

$$\Delta_j = \frac{p_0 p_1 \cdot \cdot \cdot p_{j-1}}{\bar{p}_1 \bar{p}_2 \cdot \cdot \cdot \bar{p}_j}$$

and the summation is to be extended to every possible value of j.

This, then, is a perfectly general formula that applies to all such problems, no matter what the shape of the volume element may be, and no matter whether the entrances and exits are influenced by the presence of other particles or not; for we have introduced no geometrical ideas into our discussion which would cause the shape of the volume element to invalidate it, and we have taken account of the fact that p and $\bar{p}$ might be different for different values of j.

Let us now pass to the more specific case of a gas, the molecules of which obey Maxwell's Law. We shall assume that we are dealing with an element of perfectly arbitrary shape, the volume of which is V and the superficial area A. A molecule that was within this element at the time t will have passed out at the time $t + dt$ if it lay within a distance $u\,dt$ of the surface, and possessed a velocity component u normal to it. We cannot specify the direction of the normal exactly, of course, for the element is supposed to be of purely arbitrary shape. But this is of no consequence, since we know that the chance of a molecule having a velocity component u in a given direction is the same for all directions. It is, in fact, given by (77). As for the chance of it lying within the necessary distance of the surface, that is just the ratio of the volume of the shell within which it must lie, to the total volume V. That is, it is $A\,u\,dt/V$. We have, then, for the chance of a molecule passing over the boundary during the interval dt, provided it is within the element when that interval begins, the formula

$$\frac{A\,dt}{V}\int_0^\infty u\,p^*(u)\,du = \frac{A\,dt}{2V\sqrt{a\pi}}.$$

Of course, if there are j molecules in the element to begin with, the chance of one passing out during dt is just j times as great. Hence we have

$$\bar{p}_j\,dt = \frac{j\,A\,dt}{2V\sqrt{a\pi}}. \tag{247}$$

As for the chance of a molecule passing *into* V, to do that it must lie within a distance u of the boundary on the outside if it has a normal velocity u. But by (239) and (240) the chance of a molecule of velocity U within such an element of volume is

$$\nu\left(\frac{a}{\pi}\right)^{3/2} e^{-a(u^2+v^2+w^2)}\,Au\,dt,$$

wherefore, upon integrating over every possible value of v and w (that is, of the tangential components) and over those values of u which are directed *toward* our element instead of away from it, we have

$$p\,dt = \nu A\,dt\left(\frac{a}{\pi}\right)^{3/2}\int_0^\infty du\int_{-\infty}^{\infty} dv\int_{-\infty}^{\infty} dw\,e^{-a(u^2+v^2+w^2)} = \frac{\nu A\,dt}{2\sqrt{a\pi}}. \tag{248}$$

If we now introduce these results into (246), and note that νV is just the expected number of particles in volume V, we get at once

$$\Delta_j = \frac{(\nu V)^j}{j!} = \frac{\epsilon^j}{j!}.$$

For the denominator we have, if λ is regarded as infinite,[1]

$$\sum_{j=0}^{\infty}\frac{\epsilon^j}{j!} = e^{-\epsilon},$$

whence

$$P(j) = \frac{\epsilon^j e^{-\epsilon}}{j!},$$

which is just the familiar Poisson Law.

[1] We have already tacitly assumed the number of molecules to be infinite in deriving the formula for p. For if there are only λ molecules, where λ is at all comparable with j, the presence of j of them within the element V influences the chance of another being within the shell from which new entrants come. The reader will have no difficulty in revising the formula to deal with such a case.

It may be of interest to make use of this result to indicate how large the density fluctuations are within a small portion of gas. The density of the gas is, of course, proportional to the number of molecules j. If an element contains $j = \epsilon + \delta$ molecules, therefore, instead of the expected ϵ, its density is just $j/\epsilon = 1 + \delta/\epsilon$ times the expected value.

We consider a cubical element of volume, 0.01 cm. on a side, within a gas at room temperature and atmospheric pressure. It is known that in such a gas the number of molecules per cubic centimeter is about $2.5 \cdot 10^{19}$. This is our value of ν. As $V = 10^{-6}$ it follows that the expected number of molecules within this small volume is $\epsilon = 2.5 \cdot 10^{13}$. How large, now, have we reason to expect the fluctuations in this number to be?

In our discussion of the fit of statistical data we have learned to regard the standard deviation as an indication of the spread of our distribution function, and we find from Appendix X that the standard deviation of the Poisson Formula is just $\sqrt{\epsilon}$. In our illustration this works out to be 5,000,000. Values of δ of the order of magnitude of 5,000,000 may therefore be regarded as quite usual. But a deviation from expectation of this magnitude in the number of molecules contained in the element leads to a density only 1.0000002 times its normal value.[1] So in volume elements of the size under consideration the density fluctuations are very slight indeed.

If, however, we were to consider a cube the dimensions of which were comparable to a wave-length of light we would find that the density fluctuations were appreciable, and if we were to deal with a fluid in which colloidal particles were suspended there might be considerable fluctuations in even larger elements. Such fluctuations are believed to be the cause of the optical phenomenon known as opalescence.

§ 155. *The Rapidity of Density Fluctuations*

It may appear strange that the results of the last paragraph are entirely independent of the shape of the volume element

[1] Or to the reciprocal of this value if δ is negative.

which we have under discussion, since the probability of a molecule either entering or leaving the element is proportional to the superficial area. Our result says, in fact, that if we were to immerse a cubical box, with only a pinhole in its walls, within a large flask of gas, large deviations from expectation would be just as likely as if the element were bounded only by imaginary walls through which the molecules could pass with absolute freedom. There must, however, be *some* difference between the two cases, and it is the purpose of the present section to discover what it is.

We return to the general differential equation (245) and insert the values of p and $\bar{p}$ obtained in (247) and (248). Writing $1/\alpha$ in place of $\nu A/2\sqrt{a\pi}$, so that $p_j = 1/\alpha$ and $\bar{p}_j = j/\alpha\epsilon$, we have

$$\frac{dP(j)}{dt} = \frac{1}{\alpha} P(j-1) - \frac{1}{\alpha}\left(1 + \frac{j}{\epsilon}\right) P(j) + \frac{1}{\alpha}\frac{j+1}{\epsilon} P(j+1).$$

This is the differential equation which our system must satisfy during the period when it is returning to normal after its statistical equilibrium has been somehow upset. We ought, therefore, to be able to learn from it something about the speed with which such a return to normal takes place.

The constant α occurs as a divisor of the entire right-hand side of this equation; so if we were to introduce a new unit of time defined by the relation

$$t = \alpha\tau$$

we would arrive at a differential equation (and therefore also at a solution) which was entirely independent of α. In other words, if we are dealing with *two* elements of volume, within each of which the expected number of molecules is ϵ but for which the areas are different, so that the value of α is not the same for both, and if in speaking of what goes on in the first element we measure time in terms of a unit which is α_1 seconds in length, while for the second element we use a unit which is

α_2 units long, then *in terms of these units* the one system will recover from an abnormal condition in just the same time as the other. But of course this means that the one which has the larger unit of time will actually take a larger number of seconds, in proportion to the relative magnitudes of the two α's. Remembering the definition of α, we see that this is equivalent to the statement:

Though the MAGNITUDE *of statistical density fluctuations is dependent only upon the expected density and is influenced neither by the shape of the volume element nor by the area of the surface across which migrations can take place, the* RAPIDITY *of these fluctuations varies inversely as the area of this bounding surface.*

It is the rapidity of the fluctuation, then, and not its magnitude, which is dependent upon the shape of the boundary. If we return to our specific illustration of the cubical elements of gas, one bounded by a physical container perforated by a pinhole and the other bounded only by a mathematical surface, and suppose that at some instant both are completely empty, it is evident that the mathematical one would fill up much more rapidly than the other because its effective area is so much greater. This is quite in agreement with our common sense. But if we were to make many random observations upon each of them, after they had reached the condition of statistical equilibrium, we could expect the same proportion of those observations to show a given deviation from the expected density in each case.

We still do not have a definite measure of the number of seconds required to return to equilibrium, and indeed the attempt to obtain such a definite measure is a difficult one. There are two reasons for this. In the first place, the return to equilibrium is asymptotic in character, and it is not a simple matter to specify what the "end" of the process shall be. In the second place, an exact answer to the question would require the solution of the set of equations (245), which is by no means simple.

We can, however, get some insight into the matter by noting that a solution of (245) can be obtained in the form

$$P(j) = \frac{\epsilon^j e^{-\epsilon}}{j!}\left(1 + c_0 e^{-\alpha t} + c_1 e^{-\left(1+\frac{1}{\epsilon}\right)\alpha t} + c_2 e^{-\left(1+\frac{2}{\epsilon}\right)\alpha t} + \ldots\right).$$

The complete solution will, of course, consist of as many equations of this sort as there are values of j, and the c's will be different for each of them. Moreover, these c's will depend upon the nature of the statistical upset from which the system is recovering. But the nature of the exponential terms will always be the same, no matter what value of j we may consider, nor what the statistical abnormality may have been. Hence the terms which contain the c's all die out exponentially in such a way as to reach $1/e$ times their original value in $1/\alpha$ seconds or less, and 1 per cent of their original value in a little less than five times as long. Therefore, unless the statistical upset has been of such a nature as to make the c's inordinately large, of which we can only be sure by actually solving the equations, the time of recovery may very properly be said to be a small multiple of α. This is as far as we can go without extensive computation.

§ 156. *The Schottky Effect*

In a vacuum tube the current which passes from the filament to the plate is a migration of discrete particles which possesses regularity only in a statistical sense. Either the discrete nature of the particles or the inequality in the time intervals between them would be sufficient to affect any receiving circuit that could distinguish between steady and variable currents. The magnitude of the effect might conceivably be very small, but if it were sufficiently amplified it could be made to actuate a telephone receiver. It would then manifest itself as an audible noise.

If some other fluctuation, such as an attenuated telephone message, were superposed on the original space current, it would be amplified to exactly the same degree as the statistical fluctuations and would likewise result in an audible signal.

The intelligibility of the message thus received would then depend upon the ratio of the signal amplitude to the amplitude of the statistical variations. If this ratio were too small, the noise would mask the signal and the result would be worthless for communication purposes. In other words, even with perfect amplifiers, there is a lower limit of intensity below which electrical signals cannot be detected by vacuum tubes because of the lumpiness of the space current in the first tube.

As a final illustration of fluctuation phenomena we shall consider this problem, with a view to finding how much more electrical energy would be dissipated in a measuring circuit because of these statistical fluctuations than would be the case if the stream were perfectly steady.

We shall assume:

(*a*) That the system obeys the "law of superposition," which means only that the current which flows in response to the sum of two electromotive forces simultaneously applied is the sum of the currents which would result if each force were applied separately;

(*b*) That the emitting and measuring systems are in "statistical equilibrium";

(*c*) That the instants at which electrons emerge are distributed individually and collectively at random;

(*d*) That the magnitude of the Schottky effect is defined to be the difference between the expectation of the power in the measuring device, and the power that would be dissipated if all the irregularities of the electron stream were smoothed out.

Consider a time interval T so long that the cumulative current and electromotive force due to all those electrons which were emitted before the interval began have practically vanished by the time it ends.[1] At the instant when this interval ends there is a power EI in the measuring device.

[1] It may be noted in passing that (mathematically) non-dissipative measuring devices are ruled out by this condition. Such devices are not used by the physicist, however. Even the idealized non-dissipative circuits which he deals with in theoretical studies are the limits toward which dissipative circuits approach as the dissipation is caused to vanish. Such circuits can be treated by the methods here used.

Its value is unknown, of course, but we shall find that we can compute its *expectation*.

We denote by $p(n)$ the chance of just n electrons passing during this interval T. Due to assumption (c) we can write a formula for it at once as

$$p(n) = \frac{e^{-\nu T}(\nu T)^n}{n!},$$

where ν is the rate of emission. We also denote by $P_n(EI)$ the conditional probability of a power exactly equal to EI at the end of the interval T, if just n electrons were emitted. In terms of these symbols the expected power is obviously

$$\epsilon(EI) = \Sigma\, EI\, p(n)\, P_n(EI),$$

the summation being extended to every possible value of n and EI. Since E and I are continuous variables and n is a discontinuous one, this requires the use of both integration and algebraic summation, and takes the form

$$\epsilon(EI) = \sum_0^\infty p(n) \int_0^\infty dE \int_0^\infty dI\, EI\, P_n(EI). \qquad (249)$$

We shall first consider the integral terms of this expression, taken alone. As they form a sort of "conditional expectation" we shall represent them by the symbol [1] $\epsilon_n(EI)$.

We now note that, if we begin with a group of s points, placed at random upon a line segment, and to them add s' more, also placed at random, the result is a random group of $s + s'$ points. In other words, the superposition of two random groups upon a line produces nothing else than a larger random group. In particular, we might start with a group of $n - 1$ and add to it another group of 1, both groups being placed at random. The result would then be a random group of n. Conversely, we may regard any group of n as having been formed in this way.

Let us think, then, of the n electrons which have been

[1] This can be done without danger of confusion with an nth expectation, since we have in this problem no need for expectations higher than the first.

emitted within an interval T as constituting two groups, one $n-1$ in number and the other consisting of a single electron. In addition, let us denote the electromotive forces and currents due to these sets separately by E_1, I_1 and E_2, I_2, and their respective probabilities by $P_{n-1}(E_1, I_1)$ and $P_1(E_2, I_2)$. By assumption (a) the aggregate current and electromotive force due to the superposition of the two are $I_1 + I_2$ and $E_1 + E_2$, wherefore the instantaneous power must be $(E_1 + E_2)(I_1 + I_2)$. Using these values we obtain at once the formula

$$\epsilon_n(EI) = \int dE_1 \int dE_2 \int dI_1 \int dI_2 [E_1 I_1 + E_2 I_1 + E_1 I_2 + E_2 I_2] P_{n-1}(E_1, I_1) P_1(E_2, I_2),$$

the limits of integration being, in each case, from $-\infty$ to $+\infty$.

Let us now inspect the four terms of this integral separately. So far as the first is concerned, E_2 and I_2 occur only in $P_1(E_2, I_2)$, the integral of which over all possible values of its arguments must give unity. The entire first integral therefore reduces to

$$\int dE_1 \int dI_1 \, E_1 I_1 \, P_{n-1}(E_1 I_1),$$

which is by definition $\epsilon_{n-1}(EI)$, since the subscripts serve no other purpose than to keep our variables distinct.

In a similar fashion, the fourth term reduces immediately to $\epsilon_1(EI)$.

As for the second term, it can be separated into the product of the two expressions

$$\int dE_2 \int dI_2 \, E_2 \, P_1(E_2, I_2)$$

and

$$\int dE_1 \int dI_1 \, I_1 \, P_{n-1}(E_1, I_1),$$

which are obviously the expected value of E at the end of an interval in which only one electron was emitted, and the expected value of I at the end of an interval in which $n-1$ were emitted. We denote these by $\epsilon_1(E)$ and $\epsilon_{n-1}(I)$.

Similarly the third term of the integral leads to $\epsilon_1(I)$ and $\epsilon_{n-1}(E)$.

Collecting these results we have

$$\epsilon_n(EI) = \epsilon_{n-1}(EI) + \epsilon_1(E)\,\epsilon_{n-1}(I) + \epsilon_{n-1}(E)\,\epsilon_1(I) + \epsilon_1(EI). \quad (250)$$

By virtue of assumption (*a*) it follows at once, however, that

$$\epsilon_{n-1}(E) = (n-1)\,\epsilon_1(E),$$
$$\epsilon_{n-1}(I) = (n-1)\,\epsilon_1(I);$$

whence (250) becomes

$$\epsilon_n(EI) = \epsilon_{n-1}(EI) + \epsilon_1(EI) + 2(n-1)\,\epsilon_1(E)\,\epsilon_1(I);$$

from which, by setting n equal to 2, 3, 4, ... in succession, we easily derive the general formula

$$\epsilon_n(EI) = n\,\epsilon_1(EI) + n(n-1)\,\epsilon_1(E)\,\epsilon_1(I). \quad (251)$$

The computation of $\epsilon(EI)$ is now possible. We replace the integral terms of (249) by the expression (251) to which we have succeeded in reducing them and evaluate the summation, thus arriving at the final result

$$\epsilon(EI) = (\nu T)\,\epsilon_1(EI) + (\nu T)^2\,\epsilon_1(E)\,\epsilon_1(I). \quad (252)$$

This equation actually contains the solution of our problem. All that remains is to interpret it. This we can easily do by noting that the instant at which our interval ends is itself of the nature of a random observation upon the system. Suppose, however, that we have any function of the time $f(t)$, which extends over an interval of length T. If we choose a time at random and observe this function, we are as likely to land in any element dt as any other. Hence the expectation of the value of f shown by our observation is

$$\epsilon(f) = \frac{1}{T}\int_0^T f(t)\,dt,$$

which is just the height of a rectangle the area of which is the same as the area under $f(t)$. In other words, $\epsilon(f)$ is the

value which $f(t)$ would have if all its fluctuations were smoothed out.

In the case of the functions EI, E and I in (252), where T has been chosen much longer than the physical duration of the pulse due to an electron, the integral from 0 to T includes the entire area under the curve in question. In other words, $T\epsilon_1(EI)$ is the total energy which would be dissipated by a single electron if no other were ever emitted, which we may call w_1; while $\nu T\epsilon_1(E)$ and $\nu T\epsilon_1(I)$ are the values which the current and voltage due to νT electrons would have, if all the fluctuations were smoothed out. Their product, then, is the constant power that would exist under these circumstances. We denote it by W_0, thus reaching the result

$$\epsilon(EI) = \nu w_1 + W_0.$$

According to the definition contained in (d), the difference between this expected power and W_0 is just the magnitude of the Schottky effect for which we are seeking. Thus we have,

$$S = \nu w_1.$$

Stated in words this simple expression says,

If the receiving circuit is of such a nature that the emission of a single electron would cause the dissipation of w_1 units of energy in it in the absence of all other electrons, and if electrons are being emitted at the rate of ν per unit time, the power in the circuit exceeds that which would be produced by the emission of a perfectly steady stream of electricity by the amount νw_1.

This is a remarkably simple result for such a complicated piece of reasoning; and incidentally one which fits in well with the experimental requirements of the physicist, for in order to determine the quantity w_1 he need only subject his measuring circuit to a shock such as that which an electron would give it, but on a larger scale, and measure the total amount of energy dissipated as a result. Reduced to the scale of a single electron, and multiplied by the rate of emission ν, it gives him at once the magnitude of the Schottky effect.

References for Outside Reading

I. *The Kinetic Theory:*

1. Jeans: *The Dynamical Theory of Gases.*
2. Watson: *A Treatise on the Kinetic Theory of Gases.*
3. Boltzman: *Vorlesungen über Gastheorie.*

II. *Fluctuation Phenomena in General:*

4. Fürth: *Schwankungserscheinungen in der Physik.*

III. *The Schottky Effect:*

5. Schottky: *Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern*, Annalen der Physik, Vol. 57 (1918), pp. 541–567.
6. Furth: *Die Bestimmung der Elektronenladung aus dem Schroteffekt an Glühkathodenröhren*, Physikalische Zeitschrift, Vol. 23 (1922), pp. 354–362.
7. Fry: *The Theory of the Schroteffekt*, Journal of the Franklin Institute, Vol. 199 (1925), pp. 203–220.

APPENDICES

n	$n!$	n	$n!$	n	$n!$
1	1	36	$3.719\ 9333^{41}$	71	$8.504\ 7859^{101}$
2	2	37	$1.376\ 3753^{43}$	72	$6.123\ 4458^{103}$
3	6	38	$5.230\ 2262^{44}$	73	$4.470\ 1155^{105}$
4	2.4^{1}	39	$2.039\ 7882^{46}$	74	$3.307\ 8854^{107}$
5	1.20^{2}	40	$8.159\ 1528^{47}$	75	$2.480\ 9141^{109}$
6	7.20^{2}	41	$3.345\ 2527^{49}$	76	$1.885\ 4947^{111}$
7	5.040^{3}	42	$1.405\ 0061^{51}$	77	$1.451\ 8309^{113}$
8	$4.032\ 0^{4}$	43	$6.041\ 5263^{52}$	78	$1.132\ 4281^{115}$
9	$3.628\ 80^{5}$	44	$2.658\ 2716^{54}$	79	$8.946\ 1821^{116}$
10	$3.628\ 800^{6}$	45	$1.196\ 2222^{56}$	80	$7.156\ 9457^{118}$
11	$3.991\ 6800^{7}$	46	$5.502\ 6222^{57}$	81	$5.797\ 1260^{120}$
12	$4.790\ 0160^{8}$	47	$2.586\ 2324^{59}$	82	$4.753\ 6433^{122}$
13	$6.227\ 0208^{9}$	48	$1.241\ 3916^{61}$	83	$3.945\ 5240^{124}$
14	$8.717\ 8291^{10}$	49	$6.082\ 8186^{62}$	84	$3.314\ 2401^{126}$
15	$1.307\ 6744^{12}$	50	$3.041\ 4093^{64}$	85	$2.817\ 1041^{128}$
16	$2.092\ 2790^{13}$	51	$1.551\ 1188^{66}$	86	$2.422\ 7095^{130}$
17	$3.556\ 8743^{14}$	52	$8.065\ 8175^{67}$	87	$2.107\ 7573^{132}$
18	$6.402\ 3737^{15}$	53	$4.274\ 8833^{69}$	88	$1.854\ 8264^{134}$
19	$1.216\ 4510^{17}$	54	$2.308\ 4370^{71}$	89	$1.650\ 7955^{136}$
20	$2.432\ 9020^{18}$	55	$1.269\ 6403^{73}$	90	$1.485\ 7160^{138}$
21	$5.109\ 0942^{19}$	56	$7.109\ 9859^{74}$	91	$1.352\ 0015^{140}$
22	$1.124\ 0007^{21}$	57	$4.052\ 6920^{76}$	92	$1.243\ 8414^{142}$
23	$2.585\ 2017^{22}$	58	$2.350\ 5613^{78}$	93	$1.156\ 7725^{144}$
24	$6.204\ 4840^{23}$	59	$1.386\ 8312^{80}$	94	$1.087\ 3662^{146}$
25	$1.551\ 1210^{25}$	60	$8.320\ 9871^{81}$	95	$1.032\ 9978^{148}$
26	$4.032\ 9146^{26}$	61	$5.075\ 8021^{83}$	96	$9.916\ 7793^{149}$
27	$1.088\ 8869^{28}$	62	$3.146\ 9973^{85}$	97	$9.619\ 2760^{151}$
28	$3.048\ 8834^{29}$	63	$1.982\ 6083^{87}$	98	$9.426\ 8904^{153}$
29	$8.841\ 7620^{30}$	64	$1.268\ 8693^{89}$	99	$9.332\ 6215^{155}$
30	$2.652\ 5286^{32}$	65	$8.247\ 6506^{90}$	100	$9.332\ 6215^{157}$
31	$8.222\ 8387^{33}$	66	$5.443\ 4494^{92}$	101	$9.425\ 9478^{159}$
32	$2.631\ 3084^{35}$	67	$3.647\ 1111^{94}$	102	$9.614\ 4667^{161}$
33	$8.683\ 3176^{36}$	68	$2.480\ 0355^{96}$	103	$9.902\ 9007^{163}$
34	$2.952\ 3280^{38}$	69	$1.711\ 2245^{98}$	104	$1.029\ 9017^{166}$
35	$1.033\ 3148^{40}$	70	$1.197\ 8572^{100}$	105	$1.081\ 3968^{168}$

n	$n!$	n	$n!$	n	$n!$
106	$1.146\ 2806^{170}$	141	$1.898\ 1438^{243}$	176	$1.979\ 0311^{320}$
107	$1.226\ 5202^{172}$	142	$2.695\ 3641^{245}$	177	$3.502\ 8851^{322}$
108	$1.324\ 6418^{174}$	143	$3\ 854\ 3707^{247}$	178	$6\ 235\ 1354^{324}$
109	$1.443\ 8596^{176}$	144	$5.550\ 2938^{249}$	179	$1.116\ 0892^{327}$
110	$1.588\ 2455^{178}$	145	$8.047\ 9261^{251}$	180	$2.008\ 9606^{329}$
111	$1.762\ 9526^{180}$	146	$1.174\ 9972^{254}$	181	$3.636\ 2187^{331}$
112	$1.974\ 5069^{182}$	147	$1.727\ 2459^{256}$	182	$6.617\ 9181^{333}$
113	$2.231\ 1927^{184}$	148	$2.556\ 3239^{258}$	183	$1.211\ 0790^{336}$
114	$2.543\ 5597^{186}$	149	$3.808\ 9226^{260}$	184	$2.228\ 3854^{338}$
115	$2.925\ 0937^{188}$	150	$5.713\ 3840^{262}$	185	$4.122\ 5130^{340}$
116	$3.393\ 1087^{190}$	151	$8.627\ 2098^{264}$	186	$7.667\ 8741^{342}$
117	$3.969\ 9372^{192}$	152	$1.311\ 3359^{267}$	187	$1.433\ 8925^{345}$
118	$4.684\ 5258^{194}$	153	$2.006\ 3439^{269}$	188	$2.695\ 7178^{347}$
119	$5.574\ 5858^{196}$	154	$3.089\ 7696^{271}$	189	$5.094\ 9067^{349}$
120	$6.689\ 5029^{198}$	155	$4.789\ 1429^{273}$	190	$9.680\ 3227^{351}$
121	$8.094\ 2985^{200}$	156	$7.471\ 0629^{275}$	191	$1.848\ 9416^{354}$
122	$9.875\ 0442^{202}$	157	$1.172\ 9569^{278}$	192	$3.549\ 9679^{356}$
123	$1.214\ 6304^{205}$	158	$1.853\ 2719^{280}$	193	$6.851\ 4381^{358}$
124	$1.506\ 1417^{207}$	159	$2.946\ 7023^{282}$	194	$1.329\ 1790^{361}$
125	$1.882\ 6772^{209}$	160	$4.714\ 7236^{284}$	195	$2.591\ 8990^{363}$
126	$2.372\ 1732^{211}$	161	$7.590\ 7051^{286}$	196	$5.080\ 1221^{365}$
127	$3.012\ 6600^{213}$	162	$1.229\ 6942^{289}$	197	$1.000\ 7841^{368}$
128	$3.856\ 2048^{215}$	163	$2.004\ 4016^{291}$	198	$1.981\ 5524^{370}$
129	$4.974\ 5042^{217}$	164	$3.287\ 2186^{293}$	199	$3.943\ 2893^{372}$
130	$6.466\ 8555^{219}$	165	$5.423\ 9107^{295}$	200	$7.886\ 5787^{374}$
131	$8.471\ 5807^{221}$	166	$9.003\ 6917^{297}$		
132	$1.118\ 2487^{224}$	167	$1.503\ 6165^{300}$		
133	$1.487\ 2707^{226}$	168	$2.526\ 0757^{302}$		
134	$1.992\ 9427^{228}$	169	$4.269\ 0680^{304}$		
135	$2.690\ 4727^{230}$	170	$7.257\ 4156^{306}$		
136	$3.659\ 0429^{232}$	171	$1.241\ 0181^{309}$		
137	$5.012\ 8887^{234}$	172	$2.134\ 5511^{311}$		
138	$6.917\ 7865^{236}$	173	$3.692\ 7734^{313}$		
139	$9.615\ 7232^{238}$	174	$6.425\ 4257^{315}$		
140	$1.346\ 2012^{241}$	175	$1.124\ 4495^{318}$		

$\frac{1}{2}! = \frac{1}{2}\sqrt{\pi}$; $(-\frac{1}{2})! = \sqrt{\pi}$; $(n+\frac{1}{2})! = (n+\frac{1}{2})(n-\frac{1}{2})(n-\frac{3}{2})\ldots\frac{3}{2}\cdot\frac{1}{2}\sqrt{\pi}$.

n	$\log n!$	n	$\log n!$	n	$\log n!$
1	0.000 000 0000	41	49.524 428 9249	81	120.763 212 7414
2	0.301 029 9957	42	51.147 678 2153	82	122.677 026 5938
3	0.778 151 2504	43	52.781 146 6709	83	124.596 104 6861
4	1.380 211 2417	44	54.424 599 3473	84	126.520 383 9722
5	2.079 181 2460	45	56.077 811 8611	85	128.449 802 8979
6	2.857 332 4964	46	57.740 569 6928	86	130.384 301 3492
7	3.702 430 5364	47	59.412 667 5507	87	132.323 820 6018
8	4.605 520 5234	48	61.093 908 7881	88	134.268 303 2739
9	5.559 763 0329	49	62.784 104 8681	89	136.217 693 2806
10	6.559 763 0329	50	64.483 074 8725	90	138.171 935 7900
11	7.601 155 7180	51	66.190 645 0486	91	140.130 977 1823
12	8.680 336 9641	52	67.906 648 3922	92	142.094 765 0097
13	9.794 280 3164	53	69.630 924 2618	93	144.063 247 9582
14	10.940 408 3521	54	71.363 318 0216	94	146.036 375 8118
15	12.116 499 6111	55	73.103 680 7111	95	148.014 099 4171
16	13.320 619 5938	56	74.851 868 7381	96	149.996 370 6502
17	14.551 068 5152	57	76.607 743 5938	97	151.983 142 3844
18	15.806 341 0203	58	78.371 171 5874	98	153.974 368 4601
19	17.085 094 6212	59	80.142 023 5990	99	155.970 003 6547
20	18.386 124 6169	60	81.920 174 8494	100	157.970 003 6547
21	19.708 343 9116	61	83.705 504 6844	101	159.974 325 0285
22	21.050 766 5924	62	85.497 896 3739	102	161.982 925 2003
23	22.412 494 4285	63	87.297 236 9234	103	163.995 762 4250
24	23.792 705 6702	64	89.103 416 8973	104	166.012 795 7643
25	25.190 645 6788	65	90.916 330 2540	105	168.033 985 0633
26	26.605 619 0268	66	92.735 874 1895	106	170.059 290 9286
27	28.036 982 7910	67	94.561 948 9922	107	172.088 674 7063
28	29.484 140 8223	68	96.394 457 9049	108	174.122 098 4618
29	30.946 538 8202	69	98.233 306 9957	109	176.159 524 9597
30	32.423 660 0749	70	100.078 405 0357	110	178.200 917 6449
31	33.915 021 7688	71	101.929 663 3844	111	180.246 240 6237
32	35.420 171 7471	72	103.786 995 8808	112	182.295 458 6463
33	36.938 685 6870	73	105.650 318 7410	113	184.348 537 0898
34	38.470 164 6040	74	107.519 550 4607	114	186.405 441 9411
35	40.014 232 6484	75	109.394 611 7241	115	188.466 139 7815
36	41.570 535 1491	76	111.275 425 3164	116	190.530 597 7707
37	43.138 736 8732	77	113.161 916 0415	117	192.598 783 6325
38	44.718 520 4698	78	115.054 010 6442	118	194.670 665 6398
39	46.309 585 0768	79	116.951 637 7355	119	196.746 212 6012
40	47.911 645 0682	80	118.854 727 7225	120	198.825 393 8472

Taken from the 18-place table of C. F. Degen, Havniae, 1824,

n	$\log n!$	n	$\log n!$	n	$\log n!$
121	200.908 179 2175	161	286.880 282 1167	201	377.200 084 6975
122	202.994 539 0482	162	289.089 797 1313	202	379.505 436 0669
123	205.084 444 1597	163	291.301 984 7357	203	381.812 932 1048
124	207.177 865 8448	164	293.516 828 5837	204	384.122 562 2722
125	209.274 775 8578	165	295.734 312 5279	205	386 434 316 1333
126	211.375 146 4029	166	297.954 420 6160	206	388.748 183 3537
127	213.478 950 1239	167	300.177 137 0871	207	391.064 153 6991
128	215.586 160 0935	168	302.402 446 3688	208	393.382 217 0341
129	217.696 749 8038	169	304.630 333 0735	209	395.702 363 3202
130	219.810 693 1561	170	306.860 781 9948	210	398.024 582 6149
131	221.927 964 4518	171	309.093 778 1052	211	400.348 865 0702
132	224.048 538 3830	172	311.329 306 5521	212	402.675 200 9312
133	226.172 390 0240	173	313.567 352 6553	213	405.003 580 5346
134	228.299 494 8223	174	315.807 901 9035	214	407.333 994 3080
135	230.429 828 5908	175	318.050 939 9522	215	409.666 432 7679
136	232.563 367 4992	176	320.296 452 6200	216	412.000 886 5190
137	234.700 088 0664	177	322.544 425 8864	217	414.337 346 2529
138	236.839 967 1528	178	324 794 845 8887	218	416 675 802 7465
139	238.982 981 9530	179	327.047 698 9197	219	419.016 246 8613
140	241.129 109 9887	180	329.302 971 4248	220	421 358 669 5421
141	243.278 329 1014	181	331.560 649 9997	221	423.703 061 8158
142	245.430 617 4457	182	333.820 721 3876	222	426.049 414 7903
143	247.585 953 4832	183	336.083 172 4774	223	428.397 719 6533
144	249.744 315 9753	184	338 347 990 3004	224	430 747 967 6717
145	251.905 683 9775	185	340.615 162 0288	225	433.100 150 1898
146	254.070 036 8333	186	342.884 674 9730	226	435.454 258 6289
147	256.237 354 1681	187	345.156 516 5795	227	437.810 284 4861
148	258.407 615 8835	188	347.430 674 4288	228	440.168 219 3331
149	260.580 802 1519	189	349.707 136 2330	229	442.528 054 8154
150	262.756 893 4109	190	351.985 889 8339	230	444.889 782 6515
151	264.935 870 3582	191	354.266 923 2012	231	447.253 394 6314
152	267.117 713 9462	192	356.550 224 4299	232	449.618 882 6162
153	269.302 405 3770	193	358.835 781 7389	233	451.986 238 5373
154	271.489 926 0978	194	361.123 583 4688	234	454.355 454 3947
155	273.680 257 7960	195	363.413 618 0802	235	456.726 522 2570
156	275.873 382 3943	196	365.705 874 1515	236	459.099 434 2599
157	278.069 282 0468	197	368.000 340 3777	237	461.474 182 6059
158	280.267 939 1337	198	370.297 005 5680	238	463.850 759 5630
159	282.469 336 2580	199	372.595 858 6444	239	466.229 157 4639
160	284.673 456 2407	200	374 896 888 6400	240	468.609 368 7056

n	$\log n!$	n	$\log n!$	n	$\log n!$
241	470.991 385 7482	281	567.673 298 3669	321	666.832 041 9066
242	473.375 201 1142	282	570.123 547 4752	322	669.339 897 7783
243	475.760 807 3878	283	572.575 333 9108	323	671.849 100 3006
244	478.148 197 2141	284	575.028 652 2508	324	674.359 645 3108
245	480.537 363 2985	285	577.483 497 1108	325	676.871 528 6718
246	482.928 298 4056	286	579.939 863 1439	326	679.384 746 2718
247	485.320 995 3589	287	582.397 745 0407	327	681.899 294 0245
248	487.715 447 0397	288	584.857 137 5284	328	684.415 167 8682
249	490.111 646 3868	289	587.318 035 3712	329	686.932 363 7662
250	492.509 586 3955	290	589.780 433 3691	330	689.450 877 7060
251	494.909 260 1169	291	592.244 326 3581	331	691.970 705 6998
252	497.310 660 6577	292	594.709 709 2095	332	694.491 843 7835
253	499.713 781 1789	293	597.176 576 8299	333	697.014 288 0170
254	502.118 614 8955	294	599.644 924 1603	334	699.538 034 4838
255	504.525 155 0760	295	602.114 746 1763	355	702.063 079 2909
256	506.933 395 0413	296	604.586 037 8873	336	704.589 418 5683
257	509.343 328 1646	297	607.058 794 3367	337	707.117 048 4691
258	511.754 947 8706	298	609.533 010 6007	338	709.645 965 1694
259	514.168 247 6346	299	612.008 681 7891	339	712.176 164 8676
260	516.583 220 9826	300	614.485 803 0438	340	714.707 643 7847
261	518.999 861 4900	301	616.964 369 5394	341	717.240 398 1636
262	521.418 162 7813	302	619.444 376 4823	342	719.774 424 2697
263	523.838 118 5298	303	621.925 819 1108	343	722.309 718 3897
264	526.259 722 4566	304	624.408 692 6944	344	724.846 276 8323
265	528.682 968 3306	305	626.892 992 5338	345	727.384 095 9274
266	531.107 849 9672	306	629.378 713 9603	346	729.923 172 0262
267	533.534 361 2286	307	631.865 852 3357	347	732.463 501 5010
268	535.962 496 0226	308	634.354 403 0522	348	735.005 080 7449
269	538.392 248 3026	309	636.844 361 5317	349	737.547 906 1719
270	540.823 612 0668	310	639.335 723 2255	350	740.091 974 2162
271	543.256 581 3576	311	641.828 483 6145	351	742.637 281 3327
272	545.691 150 2617	312	644.322 638 2085	352	745.183 823 9962
273	548.127 312 9087	313	646.818 182 5461	353	747.731 598 7016
274	550.565 063 4715	314	649.315 112 1942	354	750.280 601 9636
275	553.004 396 1654	315	651.813 422 7480	355	752.830 830 3166
276	555.445 305 2474	316	654.313 109 8306	356	755.382 280 3146
277	557.887 785 0165	317	656.814 169 0928	357	757.934 948 5307
278	560.331 829 8124	318	659.316 596 2128	358	760.488 831 5574
279	562.777 434 0157	319	661.820 386 8958	359	763.043 926 0060
280	565.224 592 0470	320	664.325 536 8742	360	765.600 228 5067

n	log $n!$	n	log $n!$	n	log $n!$
361	768.157 735 7086	401	871.409 558 5503	441	976.394 942 7311
362	770.716 444 2792	402	874.013 784 6034	442	979.040 365 0005
363	773.276 350 9042	403	876.619 089 6496	443	981.686 768 7267
364	775.837 452 2878	404	879.225 471 0147	444	984.334 151 6968
365	778.399 745 1523	405	881.832 926 0379	445	986.982 511 7078
366	780.963 226 2377	406	884.441 452 0715	446	989.631 846 5665
367	783.527 892 3019	407	887.051 046 4807	447	992.282 154 0896
368	786.093 740 1206	408	889.661 706 6438	448	994.933 432 1036
369	788.660 766 4868	409	892.273 429 9518	449	997.585 678 4446
370	791.228 968 2108	410	894.886 213 8085	450	1000.238 890 9584
371	793.798 342 1205	411	897.500 055 6304	451	1002.893 067 5003
372	796.368 885 0603	412	900.114 952 8464	452	1005.548 205 9351
373	798.940 593 8922	413	902.730 902 8981	453	1008.204 304 1371
374	801.513 465 4944	414	905.347 903 2392	454	1010.861 359 9900
375	804.087 496 7621	415	907.965 951 3359	455	1013.519 371 3866
376	806.662 684 6070	416	910.585 044 6665	456	1016.178 336 2293
377	809.239 025 9572	417	913.205 180 7215	457	1018.838 252 4293
378	811.816 517 7571	418	915.826 357 0033	458	1021.499 117 9074
379	814.395 156 9670	419	918.448 571 0263	459	1024.160 930 5929
380	816.974 940 5636	420	921.071 820 3167	460	1026.823 688 4246
381	819.555 865 5393	421	923.696 102 4125	461	1029.487 389 3500
382	822.137 928 9022	422	926.321 414 8635	462	1032.152 031 3255
383	824.721 127 6762	423	928.947 755 2308	463	1034.817 612 3165
384	827.305 458 9006	424	931.575 121 0874	464	1037.484 130 2971
385	829.890 919 6301	425	934.203 510 0175	465	1040.151 583 2500
386	832.477 506 9347	426	936.832 919 6166	466	1042.819 969 1667
387	835.065 217 8998	427	939.463 347 4916	467	1045.489 286 0472
388	837.654 049 6254	428	942.094 791 2606	468	1048.159 531 9003
389	840.243 999 2267	429	944.727 248 5528	469	1050.830 704 7430
390	842.835 063 8337	430	947.360 717 0084	470	1053.502 802 6010
391	845.427 240 5911	431	949.995 194 2785	471	1056.175 823 5081
392	848.020 526 6581	432	952.630 678 0254	472	1058.849 765 5067
393	850.614 919 2085	433	955.267 165 9217	473	1061.524 626 6475
394	853.210 415 4303	434	957.904 655 6512	474	1064.200 404 9891
395	855.807 012 5259	435	960.543 144 9082	475	1066.877 098 5988
396	858.404 707 7119	436	963.182 631 3974	476	1069.554 705 5515
397	861.003 498 2186	437	965.823 112 8344	477	1072.233 223 9305
398	863.603 381 2907	438	968.464 586 9449	478	1074.912 651 8271
399	866.204 354 1864	439	971.107 051 4652	479	1077.592 987 3405
400	868.806 414 1777	440	973.750 504 1416	480	1080.274 228 5779

n	$\log n!$	n	$\log n!$	n	$\log n!$
481	1082.956 373 6543	521	1190.962 607 0282	561	1300.302 596 6937
482	1085.639 420 6925	522	1193.680 277 5312	562	1303.052 333 0093
483	1088.323 367 8233	523	1196.398 779 2200	563	1305.802 841 4042
484	1091.008 213 1849	524	1199.118 110 5070	564	1308.554 120 5081
485	1093.693 954 9235	525	1201.838 269 8104	565	1311.306 168 9560
486	1096.380 591 1928	526	1204.559 255 5546	566	1314.058 985 3871
487	1099.068 120 1540	527	1207.281 066 1698	567	1316.812 568 4460
488	1101.756 539 9760	528	1210.003 700 0923	568	1319.566 916 7818
489	1104.445 848 8351	529	1212.727 155 7644	569	1322.322 029 0481
490	1107.136 044 9152	530	1215.451 431 6340	570	1325.077 903 9038
491	1109.827 126 4073	531	1218.176 526 1550	571	1327.834 540 0121
492	1112.519 091 5101	532	1220.902 437 7873	572	1330.591 936 0409
493	1115.211 938 4293	533	1223.629 164 9964	573	1333.350 090 6628
494	1117.905 665 3783	534	1226.356 706 2534	574	1336.109 002 5552
495	1120.600 270 5772	535	1229.085 060 0354	575	1338.868 670 3999
496	1123.295 752 2537	536	1231.814 224 8251	576	1341.629 092 8833
497	1125.992 108 6424	537	1234.544 199 1108	577	1344.390 268 6965
498	1128.689 337 9852	538	1237.274 981 3865	578	1347.152 196 5349
499	1131.387 438 5308	539	1240.006 570 1517	579	1349.914 875 0986
500	1134.086 408 5351	540	1242.738 963 9115	580	1352.678 303 0922
501	1136.786 246 2610	541	1245.472 161 1766	581	1355.442 479 2246
502	1139.486 949 9781	542	1248.206 160 4631	582	1358.207 402 2092
503	1142.188 517 9632	543	1250.940 960 2927	583	1360.973 070 7640
504	1144.890 948 4996	544	1253.676 559 1924	584	1363.739 483 6111
505	1147.594 239 8778	545	1256.412 955 6947	585	1366.506 639 4772
506	1150.298 390 3946	546	1259.150 148 3374	586	1369.274 537 0932
507	1153.003 398 3539	547	1261.888 135 6637	587	1372.043 175 1945
508	1155.709 262 0662	548	1264.626 916 2222	588	1374.812 552 5205
509	1158.415 979 8486	549	1267.366 488 5667	589	1377.582 667 8153
510	1161.123 550 0247	550	1270.106 851 2562	590	1380.353 519 8270
511	1163.831 970 9248	551	1272.848 002 8550	591	1383.125 107 3079
512	1166.541 240 8858	552	1275.589 941 9327	592	1385.897 429 0146
513	1169.251 358 2509	553	1278.332 667 0640	593	1388.670 483 7079
514	1171.962 321 3699	554	1281.076 176 8288	594	1391.444 270 1529
515	1174.674 128 5989	555	1283.820 469 8119	595	1394.218 787 1186
516	1177.386 778 3005	556	1286.565 544 6035	596	1396.994 033 3784
517	1180.100 268 8436	557	1289.311 399 7987	597	1399.770 007 7095
518	1182.814 598 6034	558	1292.058 033 9976	598	1402.546 708 8935
519	1185.529 765 9612	559	1294.805 445 8055	599	1405.324 135 7159
520	1188.245 769 3049	560	1297.553 633 8325	600	1408.102 286 9663

n	log $n!$	n	log $n!$	n	log $n!$
601	1410.881 161 4383	641	1522.615 809 4311	681	1635.434 357 6708
602	1413.660 757 9295	642	1525.423 344 4591	682	1638.268 142 0455
603	1416.441 075 2417	643	1528.231 555 4320	683	1641.102 562 7492
604	1419.222 112 1803	644	1531.040 441 2994	684	1643.937 618 8509
605	1422.003 867 5550	645	1533.850 001 0140	685	1646.773 309 4224
606	1424.786 340 1791	646	1536.660 233 5320	686	1649.609 633 5381
607	1427.569 528 8702	647	1539.471 137 8127	687	1652.446 590 2752
608	1430.353 432 4495	648	1542.282 712 8186	688	1655.284 178 7134
609	1433.138 049 7421	649	1545.094 957 5154	689	1658.122 397 9353
610	1435.923 379 5771	650	1547.907 870 8720	690	1660.961 247 0260
611	1438.709 420 7874	651	1550.721 451 8606	691	1663.800 725 0734
612	1441.496 172 2095	652	1553.535 699 4563	692	1666.640 831 1679
613	1444.283 632 6840	653	1556.350 612 6376	693	1669.481 564 4025
614	1447.071 801 0552	654	1559.166 190 3859	694	1672.322 923 8729
615	1449.860 676 1709	655	1561.982 431 6859	695	1675.164 908 6775
616	1452.650 256 8831	656	1564.799 335 5253	696	1678.007 517 9171
617	1455.440 542 0471	657	1567.616 900 8948	697	1680.850 750 6952
618	1458.231 530 5222	658	1570.435 126 7885	698	1683.694 606 1179
619	1461.023 221 1712	659	1573.254 012 2031	699	1686.539 083 2936
620	1463.815 612 8607	660	1576.073 556 1386	700	1689.384 181 3336
621	1466.608 704 4609	661	1578.893 757 5031	701	1692.229 899 3516
622	1469.402 494 8456	662	1581.714 615 5875	702	1695.076 236 4637
623	1472.196 982 8923	663	1584.536 129 1159	703	1697.923 191 7887
624	1474.992 167 4819	664	1587.358 297 1953	704	1700.770 764 4479
625	1477.788 047 4993	665	1590.181 118 8406	705	1703.618 953 5649
626	1480.584 621 8325	666	1593.004 593 0698	706	1706.467 758 2659
627	1483.381 889 3733	667	1595.828 718 9037	707	1709.317 177 6797
628	1486.179 849 0171	668	1598.653 495 3662	708	1712.167 210 9374
629	1488.978 499 6625	669	1601.478 921 4839	709	1715.017 857 1726
630	1491.777 840 2120	670	1604.304 996 2866	710	1717.869 115 5213
631	1494.577 869 5712	671	1607.131 718 8068	711	1720.720 985 1220
632	1497.378 586 6495	672	1609.959 088 0798	712	1723.573 465 1157
633	1500.179 990 3595	673	1612.787 103 1441	713	1726.426 554 6455
634	1502.982 079 6174	674	1615.615 763 0406	714	1729.280 252 8573
635	1505.784 853 3427	675	1618.445 066 8134	715	1732.134 558 8991
636	1508.588 310 4583	676	1621.275 013 5094	716	1734.989 471 9214
637	1511.392 449 8907	677	1624.105 602 1781	717	1737.844 991 0771
638	1514.197 270 5694	678	1626.936 831 8719	718	1740.701 115 5213
639	1517.002 771 4275	679	1629.768 701 6462	719	1743.557 844 4117
640	1519.808 951 4015	680	1632.601 210 5589	720	1746.415 176 9081

n	log $n!$	n	log $n!$	n	log $n!$
721	1749.273 112 1728	761	1864.075 452 8940	801	1979.790 716 7537
722	1752.131 649 3704	762	1866.957 407 8654	802	1982.694 891 1220
723	1754.990 787 6677	763	1869.839 932 4033	803	1985.599 606 6673
724	1757.850 526 2339	764	1872.723 025 7619	804	1988.504 862 7160
725	1760.710 864 2405	765	1875.606 687 1971	805	1991.410 658 5964
726	1763.571 800 8612	766	1878.490 915 9667	806	1994.316 993 6382
727	1766.433 335 2720	767	1881.375 711 3306	807	1997.223 867 1729
728	1769.295 466 6514	768	1884.261 072 5507	808	2000.131 278 5337
729	1772.158 194 1797	769	1887.146 998 8905	809	2003.039 227 0553
730	1775.021 517 0398	770	1890.033 489 6156	810	2005.947 712 0742
731	1777.885 434 4167	771	1892.920 543 9937	811	2008.856 732 9284
732	1780.749 945 4978	772	1895.808 161 2940	812	2011.766 288 9576
733	1783.615 049 4724	733	1898.696 340 7879	813	2014.676 379 5032
734	1786.480 745 5324	774	1901.585 081 7486	814	2017.587 003 9081
735	1789.347 032 8714	775	1904.474 383 4511	815	2020.498 161 5169
736	1792.213 910 6858	776	1907.364 245 1724	816	2023.409 851 6756
737	1795.081 378 1736	777	1910.254 666 1912	817	2026.322 073 7322
738	1797.949 434 5355	778	1913.145 645 7882	818	2029.234 827 0358
739	1800.818 078 9739	779	1916.037 183 2459	819	2032.148 110 9376
740	1803.687 310 6936	780	1918.929 277 8485	820	2035.061 924 7900
741	1806.557 128 9016	781	1921.821 928 8824	821	2037.976 267 9471
742	1809.427 532 8069	782	1924.715 135 6355	822	2040.891 139 7646
743	1812.298 521 6206	783	1927.608 897 3975	823	2043.806 539 5998
744	1815.170 094 5562	784	1930.503 213 4602	824	2046.722 466 8115
745	1818.042 250 8289	785	1933.398 083 1170	825	2049.638 920 7601
746	1820.914 989 6564	786	1936.293 505 6630	826	2052.555 900 8074
747	1823.788 310 2582	787	1939.189 480 3954	827	2055.473 406 3170
748	1826.662 211 8561	788	1942.086 006 6129	828	2058.391 436 6537
749	1829.536 693 6738	789	1944.983 083 6161	829	2061.309 991 1843
750	1832.411 754 9372	790	1947.880 710 7074	830	2064.229 069 2767
751	1835.287 394 8742	791	1950.778 887 1909	831	2067.148 670 3005
752	1838.163 612 7147	792	1953.677 612 3724	832	2070.068 793 6267
753	1841.040 407 6909	793	1956.576 885 5598	833	2072.989 438 6282
754	1843.917 779 0368	794	1959.476 706 0622	834	2075.910 604 6788
755	1846.795 725 9884	795	1962.377 073 1908	835	2078.832 291 1543
756	1849.674 247 7839	796	1965.277 986 2586	836	2081.754 497 4317
757	1852.553 343 6634	797	1968.179 444 5800	837	2084.677 222 8897
758	1855.433 012 8691	798	1971.081 447 4713	838	2087.600 466 9083
759	1858.313 254 6450	799	1973.983 994 2506	839	2090.524 228 8692
760	1861.194 068 2373	800	1976.887 084 2376	840	2093.448 508 1552

n	$\log n!$	n	$\log n!$	n	$\log n!$
841	2096.373 304 1510	881	2213.781 955 7223	921	2331.979 160 6232
842	2099.298 616 2425	882	2216.727 424 3075	922	2334.943 891 5443
843	2102.224 443 8171	883	2219.673 385 0110	923	2337.909 093 2453
844	2105.150 786 2638	884	2222.619 837 2760	924	2340.874 765 2165
845	2108.077 642 9727	885	2225.566 780 5467	925	2343.840 906 9493
846	2111.005 013 3358	886	2228.514 214 2686	926	2346.807 517 9360
847	2113.932 896 7461	887	2231.462 137 8885	927	2349.774 597 6701
848	2116.861 292 5984	888	2234.410 550 8542	928	2352.742 145 6463
849	2119.790 200 2886	889	2237.359 452 6152	929	2355.710 161 3603
850	2122.719 619 2143	890	2240.308 842 6219	930	2358.678 644 3089
851	2125.649 548 7744	891	2243.258 720 3259	931	2361.647 593 9898
852	2128.579 988 3692	892	2246.209 085 1803	932	2364.617 009 9022
853	2131.510 937 4003	893	2249.159 936 6392	933	2367.586 891 5459
854	2134.442 395 2710	894	2252.111 274 1580	934	2370.557 238 4222
855	2137.374 361 3857	895	2255.063 097 1933	935	2373.528 050 0331
856	2140.306 835 1504	896	2258.015 405 2029	936	2376.499 325 8818
857	2143.239 815 9723	897	2260.968 197 6460	937	2379.471 065 4727
858	2146.173 303 2602	898	2263.921 473 9826	938	2382.443 268 3111
859	2149.107 296 4240	899	2266.875 233 6744	939	2385.415 933 9033
860	2152.041 794 8753	900	2269.829 476 1838	940	2388.389 061 7569
861	2154.976 798 0267	901	2272.784 200 9748	941	2391.362 651 3803
862	2157.912 305 2925	902	2275.739 407 5123	942	2394.336 702 2831
863	2160.848 316 0883	903	2278.695 095 2626	943	2397.311 213 9759
864	2163.784 829 8307	904	2281.651 263 6931	944	2400.286 185 9702
865	2166.721 845 9382	905	2284.607 912 2723	945	2403.261 617 7787
866	2169.659 363 8302	906	2287.565 040 4700	946	2406.237 508 9151
867	2172.597 382 9277	907	2290.522 647 7571	947	2409.213 858 8941
868	2175.535 902 6529	908	2293.480 733 6056	948	2412.190 667 2314
869	2178.474 922 4293	909	2296.439 297 4888	949	2415.167 933 4439
870	2181.414 441 6819	910	2299.398 338 8811	950	2418.145 657 0491
871	2184.354 459 8370	911	2302.357 857 2581	951	2421.123 837 5661
872	2187.294 976 3219	912	2305.317 852 0964	952	2424.102 474 5145
873	2190.235 990 5656	913	2308.278 322 8740	953	2427.081 567 4151
874	2193.177 501 9982	914	2311.239 269 0697	954	2430.061 115 7898
875	2196.119 510 0512	915	2314.200 690 1638	955	2433.041 119 1614
876	2199.062 014 1574	916	2317.162 585 6374	956	2436.021 577 0537
877	2202.005 013 7508	917	2320.124 954 9731	957	2439.002 488 9914
878	2204.948 508 2667	918	2323.087 797 6543	958	2441.983 854 5005
879	2207.892 497 1418	919	2326.051 113 1657	959	2444.965 673 1077
880	2210.836 979 8139	920	2329.014 900 9930	960	2447.947 944 3407

n	$\log n!$	n	$\log n!$	n	$\log n!$
961	2450.930 667 7284	1001	2570.605 078 2996	1041	2690.973 503 8239
962	2453 913 842 8004	1002	2573 605 946 0211	1042	2693.991 371 5429
963	2456.897 469 0876	1003	2576.607 246 9542	1043	2697.009 655 8513
964	2459.881 546 1215	1004	2579.608 980 6670	1044	2700.028 356 3500
965	2462.866 073 4348	1005	2582.611 146 7287	1045	2703.047 472 6404
966	2465.851 050 5612	1006	2585.613 744 7094	1046	2706.067 004 3250
967	2468.836 477 0353	1007	2588.616 774 1800	1047	2709.086 951 0066
968	2471.822 352 3926	1008	2591.620 234 7121	1048	2712.107 312 2893
969	2474.808 676 1697	1009	2594.624 125 8783	1049	2715.128 087 7775
970	2477.795 447 9039	1010	2597.628 447 2521	1050	2718.149 277 0765
971	2480.782 667 1338	1011	2600.633 198 4077	1051	2721.170 879 7926
972	2483.770 333 3988	1012	2603.638 378 9202	1052	2724.192 895 5324
973	2486.758 446 2390	1013	2606.643 988 3656	1053	2727.215 323 9036
974	2489.747 005 1959	1014	2609.650 026 3206	1054	2730.238 164 5145
975	2492.736 009 8116	1015	2612.656 492 3628	1055	2733.261 416 9741
976	2495.725 459 6293	1016	2615.663 386 0708	1056	2736.285 080 8923
977	2498.715 354 1930	1017	2618.670 707 0237	1057	2739.309 155 8796
978	2501.705 693 0478	1018	2621.678 454 8017	1058	2742.333 641 5473
979	2504.696 475 7396	1019	2624.686 628 9857	1059	2745.358 537 5074
980	2507.687 701 8153	1020	2627.695 229 1575	1060	2748.383 843 3727
981	2510.679 370 8227	1021	2630.704 254 8996	1061	2751.409 558 7566
982	2513.671 482 3105	1022	2633.713 705 7954	1062	2754.435 683 2733
983	2516.664 035 8283	1023	2636.723 581 4291	1063	2757.462 216 5378
984	2519.657 030 9267	1024	2639.733 881 3857	1064	2760.489 158 1658
985	2522.650 467 1572	1025	2642.744 605 2511	1065	2763.516 507 7736
986	2525.644 344 0722	1026	2645.755 752 6119	1066	2766.544 264 9783
987	2528.638 661 2248	1027	2648.767 323 0555	1067	2769.572 429 3977
988	2531.633 418 1694	1028	2651.779 316 1701	1068	2772.601 000 6504
989	2534.628 614 4610	1029	2654.791 731 5449	1069	2775.629 978 3556
990	2537.624 249 6556	1030	2657.804 568 7696	1070	2778.659 362 1333
991	2540.620 323 3101	1031	2660.817 827 4349	1071	2781.689 151 6041
992	2543.616 834 9822	1032	2663.831 507 1322	1072	2784.719 346 3895
993	2546.613 784 2307	1033	2666.845 607 4537	1073	2787.749 946 1114
994	2549.611 170 6151	1034	2669.860 127 9925	1074	2790.780 950 3928
995	2552.608 993 6959	1035	2672.875 068 3422	1075	2793.812 358 8570
996	2555.607 253 0343	1036	2675.890 428 0977	1076	2796.844 171 1284
997	2558.605 948 1926	1037	2678.906 206 8540	1077	2799.876 386 8317
998	2561.605 078 7339	1038	2681.922 404 2076	1078	2802.909 005 5925
999	2564.604 644 2221	1039	2684.939 019 7551	1079	2805.942 027 0372
1000	2567.604 644 2221	1040	2687.956 053 0944	1080	2808.975 450 7927

n	$\log n!$	n	$\log n!$	n	$\log n!$
1081	2812.009 276 4866	1121	2933.687 702 5306	1161	3055.985 850 8433
1082	2815.043 503 7474	1122	2936.737 695 3876	1162	3059.051 056 9714
1083	2818.078 132 2040	1123	2939.788 075 1438	1163	3062.116 636 6861
1084	2821.113 161 4862	1124	2942.838 841 4551	1164	3065.182 589 6664
1085	2824.148 591 2244	1125	2945.889 993 9775	1165	3068.248 915 5918
1086	2827.184 421 0407	1126	2948.941 532 3680	1166	3071.315 614 1422
1087	2830.220 650 5938	1127	2951.993 456 2841	1167	3074.382 684 9982
1088	2833.257 279 4801	1128	2955.045 765 3837	1168	3077.450 127 8416
1089	2836.294 307 3689	1129	2958.098 459 3256	1169	3080.517 942 3522
1090	2839.331 733 8668	1130	2961.151 537 7691	1170	3083.586 128 2139
1091	2842.369 558 6174	1131	2964.205 000 3741	1171	3086.654 685 1090
1092	2845.407 781 2558	1132	2967.258 846 8000	1172	3089.723 612 7207
1093	2848.446 401 4177	1133	2970.313 076 7108	1173	3092.792 910 7328
1094	2851.485 418 7307	1134	2973.367 689 7653	1174	3095.862 578 8297
1095	2854.524 832 8580	1135	2976.422 685 6200	1175	3098.932 616 6963
1096	2857.564 643 4131	1136	2979.478 063 9582	1176	3102.003 024 0180
1097	2860.604 850 0400	1137	2982.533 824 4220	1177	3105.073 800 4809
1098	2863.645 452 3807	1138	2985.589 966 6850	1178	3108.144 945 7713
1099	2866.686 450 0732	1139	2988.646 490 4001	1179	3111.216 459 5764
1100	2869.727 842 7583	1140	2991.703 395 2604	1180	3114.288 341 5837
1101	2872.769 630 0773	1141	2994.760 680 9048	1181	3117.360 591 4813
1102	2875.811 811 8118	1142	2997.818 347 0087	1182	3120.433 208 9579
1103	2878.854 387 1843	1143	3000.876 393 2391	1183	3123.506 193 7025
1104	2881.897 356 2576	1144	3003.934 819 2036	1184	3126.579 545 4049
1105	2884.940 718 5357	1145	3006.993 624 7502	1185	3129.653 263 7553
1106	2887.984 473 6626	1146	3010.052 809 3079	1186	3132.727 348 4443
1107	2891.028 621 2835	1147	3013.112 372 7858	1187	3135.801 799 1632
1108	2894.073 161 0430	1148	3016.172 314 6738	1188	3138.876 615 6039
1109	2897.118 093 5901	1149	3019.232 634 7025	1189	3141.951 797 4585
1110	2900.163 415 5088	1150	3022.293 332 5429	1190	3145.027 344 4199
1111	2903.209 129 6278	1151	3025.354 407 8005	1191	3148.103 256 1814
1112	2906.255 234 4150	1152	3028.415 860 3456	1192	3151.179 532 4368
1113	2909.301 729 5794	1153	3031.477 689 0520	1193	3154.256 172 8805
1114	2912.348 614 7702	1154	3034.539 895 4617	1194	3157.333 177 2072
1115	2915.395 889 6376	1155	3037.602 477 4459	1195	3160.410 545 1125
1116	2918.443 553 8322	1156	3040.665 435 2800	1196	3163.488 276 2922
1117	2921.491 607 0053	1157	3043.728 768 6300	1197	3166.566 370 4426
1118	2924.540 048 8080	1158	3046.792 477 1984	1198	3169.644 827 2606
1119	2927.588 878 8954	1159	3049.856 560 6343	1199	3172.723 646 4437
1120	2930.638 096 9181	1160	3052.921 018 6230	1200	3175.802 827 6898

n	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$
0	1	1	1	1	1	1	1	1	1	1	1
1		1	2	3	4	5	6	7	8	9	10
2			1	3	6	10	15	21	28	36	45
3				1	4	10	20	35	56	84	120
4					1	5	15	35	70	126	210
5						1	6	21	56	126	252
6							1	7	28	84	210
7								1	8	36	120
8									1	9	45
9										1	10
10											1

n	$m=11$	$m=12$	$m=13$	$m=14$	$m=15$	$m=16$	$m=17$	$m=18$	$m=19$
0	1	1	1	1	1	1	1	1	1
1	11	12	13	14	15	16	17	18	19
2	55	66	78	91	105	120	136	153	171
3	165	220	286	364	455	560	680	816	969
4	330	495	715	1001	1365	1820	2380	3060	3876
5	462	792	1287	2002	3003	4368	6188	8568	11628
6	462	924	1716	3003	5005	8008	12376	18564	27132
7			1716	3432	6435	11440	19448	31824	50388
8					6435	12870	24310	43758	75582
9							24310	48620	92378
10									92378

n	$m=20$	$m=21$	$m=22$	$m=23$	$m=24$	$m=25$
1	2.0^{1}	2.1^{1}	2.2^{1}	2.3^{1}	2.4^{1}	2.5^{1}
2	1.90^{2}	2.10^{2}	2.31^{2}	2.53^{2}	2.76^{2}	3.00^{2}
3	1.140^{3}	1.330^{3}	1.540^{3}	1.771^{3}	2.024^{3}	2.300^{3}
4	4.845	5.985	7.315	8.855	1.0626^{4}	1.2650^{4}
5	1.5504^{4}	2.0349^{4}	2.6334^{4}	3.3649^{4}	4.2504	5.3130

Since C_0^m is equal to unity for every value of m it is not tabulated beyond $m = 20$.

n	$m = 20$	$m = 21$	$m = 22$	$m = 23$	$m = 24$	$m = 25$
6	3.876 0^{4}	5.426 4^{4}	7.461 3^{4}	1.009 47^{5}	1.345 96^{5}	1.771 00^{5}
7	7.752 0	1.162 80^{5}	1.705 44^{5}	2.451 57	3.461 04	4.807 00
8	1.259 70^{5}	2.034 90	3.197 70	4.903 14	7.354 71	1.081 575^{6}
9	1.679 60	2.939 30	4.974 20	8.171 90	1.307 504^{6}	2.042 975
10	1.847 56	3.527 16	6.466 46	1.144 066^{6}	1.961 256	3.268 760
11		3.527 16	7.054 32	1.352 078	2.496 144	4.457 400
12				1.352 078	2.704 156	5.200 300
13						5.200 300

n	$m = 26$	$m = 27$	$m = 28$	$m = 29$	$m = 30$
1	2.6^{1}	2.7^{1}	2.8^{1}	2.9^{1}	3.0^{1}
2	3.25^{2}	3.51^{2}	3.78^{2}	4.06^{2}	4.35^{2}
3	2.600^{3}	2.925^{3}	3.276^{3}	3.654^{3}	4.060^{3}
4	1.495 0^{4}	1.755 0^{4}	2.047 5^{4}	2.375 1^{4}	2.740 5^{4}
5	6.578 0	8.073 0	9.828 0	1.187 55^{5}	1.425 06^{5}
6	2.302 30^{5}	2.960 10^{5}	3.767 40^{5}	4.750 20	5.937 75
7	6.578 00	8.880 30	1.184 040^{6}	1.560 780^{6}	2.035 800^{6}
8	1.562 275^{6}	2.220 075^{6}	3.108 105	4.292 145	5.852 925
9	3.124 550	4.686 825	6.906 900	1.001 5005^{7}	1.430 7150^{7}
10	5.311 735	8.436 285	1.312 3110^{7}	2.003 0010	3.004 5015
11	7.726 160	1.303 7895^{7}	2.147 4180	3.459 7290	5.462 7300
12	9.657 700	1.738 3860	3.042 1755	5.189 5935	8.649 3225
13	1.040 0600^{7}	2.005 8300	3.744 2160	6.786 3915	1.197 5985^{8}
14		2.005 8300	4.011 6600	7.755 8760	1.454 2268
15				7.755 8760	1.551 1752

n	$m = 31$	$m = 32$	$m = 33$	$m = 34$	$m = 35$
1	3.1^{1}	3.2^{1}	3.3^{1}	3.4^{1}	3.5^{1}
2	4.65^{2}	4.96^{2}	5.28^{2}	5.61^{2}	5.95^{2}
3	4.495^{3}	4.960^{3}	5.456^{3}	5.984^{3}	6.545^{3}
4	3.146 5^{4}	3.596 0^{4}	4.092 0^{4}	4.637 6^{4}	5.236 0^{4}
5	1.699 11^{5}	2.013 76^{5}	2.373 36^{5}	2.782 56^{5}	3.246 32^{5}
6	7.362 81	9.061 92	1.107 568^{6}	1.344 904^{6}	1.623 160^{6}
7	2.629 575^{6}	3.365 856^{6}	4.272 048	5.379 616	6.724 520
8	7.888 725	1.051 8300^{7}	1.388 4156^{7}	1.815 6204^{7}	2.353 5820^{7}
9	2.016 0075^{7}	2.804 8800	3.856 7100	5.245 1256	7.060 7460
10	4.435 2165	6.451 2240	9.256 1040	1.311 2814^{8}	1.835 7940^{8}

n	$m = 31$	$m = 32$	$m = 33$	$m = 34$	$m = 35$
11	8.467 2315 ^{7}	1.290 2448 ^{8}	1.935 3672 ^{8}	2.860 9776 ^{8}	4.172 2590 ^{8}
12	1.411 2053 ^{8}	2.257 9284	3.548 1732	5.483 5404	8.344 5180
13	2.062 5308	3.473 7360	5.731 6644	9.279 8376	1.476 3378 ^{9}
14	2.651 8253	4.714 3560	8.188 0920	1.391 9756 ^{9}	2.319 9594
15	3.005 4020	5.657 2272	1.037 1583 ^{9}	1.855 9675	3.247 9432
16	3.005 4020	6.010 8039	1.166 8031	2.203 9644	4.059 9290
			1.166 8031	2.333 6062	4.537 5677
					4.537 5677

n	$m = 36$	$m = 37$	$m = 38$	$m = 39$	$m = 40$
1	3.6 ^{1}	3.7 ^{1}	3.8 ^{1}	3.9 ^{1}	4.0 ^{1}
2	6.30 ^{2}	6.66 ^{2}	7.03 ^{2}	7.41 ^{2}	7.80 ^{2}
3	7.140 ^{3}	7.770 ^{3}	8.436 ^{3}	9.139 ^{3}	9.880 ^{3}
4	5.890 5 ^{4}	6.604 5 ^{4}	7.381 5 ^{4}	8.225 1 ^{4}	9.139 0 ^{4}
5	3.769 92 ^{5}	4.358 97 ^{5}	5.019 42 ^{5}	5.757 57 ^{5}	6.580 08 ^{5}
6	1.947 792 ^{6}	2.324 784 ^{6}	2.760 681 ^{6}	3.262 623 ^{6}	3.838 380 ^{6}
7	8.347 680	1.029 5472 ^{7}	1.262 0256 ^{7}	1.538 0937 ^{7}	1.864 3560 ^{7}
8	3.026 0340 ^{7}	3.860 8020	4.890 3492	6.152 3748	7.690 4685
9	9.414 3280	1.244 0362 ^{8}	1.630 1164 ^{8}	2.119 1513 ^{8}	2.734 3888 ^{8}
10	2.541 8686 ^{8}	3.483 3014	4.727 3376	6.357 4540	8.476 6053
11	6.008 0530	8.549 9215	1.203 3223 ^{9}	1.676 0560 ^{9}	2.311 8014 ^{9}
12	1.251 6777 ^{9}	1.852 4830 ^{9}	2.707 4751	3.910 7974	5.586 8535
13	2.310 7896	3.562 4673	5.414 9503	8.122 4254	1.203 3223 ^{10}
14	3.796 2972	6.107 0868	9.669 5541	1.508 4504 ^{10}	2.320 6930
15	5.567 9026	9.364 1998	1.547 1287 ^{10}	2.514 0841	4.022 5345
16	7.307 8721	1.287 5775 ^{10}	2.223 9974	3.771 1261	6.285 2102
17	8.597 4966	1.590 5369	2.878 1143	5.102 1118	8.873 2379
18	9.075 1353	1.767 2632	3.357 8001	6.235 9144	1.133 8026 ^{11}
19		1.767 2632	3.534 5264	6.892 3264	1.312 8241
20				6.892 3264	1.378 4653

n	$m = 41$	$m = 42$	$m = 43$	$m = 44$	$m = 45$
1	4.1 ^{1}	4.2 ^{1}	4.3 ^{1}	4.4 ^{1}	4.5 ^{1}
2	8.20 ^{2}	8.61 ^{2}	9.03 ^{2}	9.46 ^{2}	9.90 ^{2}
3	1.066 0 ^{4}	1.148 0 ^{4}	1.234 1 ^{4}	1.324 4 ^{4}	1.419 0 ^{4}
4	1.012 70 ^{5}	1.119 30 ^{5}	1.234 10 ^{5}	1.357 51 ^{5}	1.489 95 ^{5}
5	7.493 98	8.506 68	9.625 98	1.086 008 ^{6}	1.221 759 ^{6}
6	4.496 388 ^{6}	5.245 786 ^{6}	6.096 454 ^{6}	7.059 052	8.145 060
7	2.248 1940 ^{7}	2.697 8328 ^{7}	3.222 4114 ^{7}	3.832 0568 ^{7}	4.537 9620 ^{7}
8	9.554 8245	1.180 3019 ^{8}	1.450 0851 ^{8}	1.772 3263 ^{8}	2.155 5320 ^{8}
9	3.503 4357 ^{8}	4.458 9181	5.639 2200	7.089 3051	8.861 6314
10	1.121 0994 ^{9}	1.471 4430 ^{9}	1.917 3348 ^{9}	2.481 2568 ^{9}	3.190 1873 ^{9}

n	$m = 41$	$m = 42$	$m = 43$	$m = 44$	$m = 45$
11	3.159 4620 ^{9}	4.280 5614 ^{9}	5.752 0043 ^{9}	7 669 3391 ^{9}	1.015 0596 ^{10}
12	7.898 6549	1.105 8117 ^{10}	1.533 8678 ^{10}	2.109 0683 ^{10}	2.876 0022
13	1.762 0076 ^{10}	2.551 8731	3.657 6848	5.191 5526	7.300 6209
14	3.524 0153	5.286 0229	7.837 8960	1.149 5581 ^{11}	1.668 7133 ^{11}
15	6.343 2275	9.867 2428	1.515 3266 ^{11}	2.299 1162	3.448 6743
16	1.030 7745 ^{11}	1.665 0972 ^{11}	2.651 8215	4.167 1481	6.466 2642
17	1.515 8448	2.546 6193	4.211 7165	6.863 5380	1.103 0686 ^{12}
18	2.021 1264	3.536 9712	6.083 5905	1.029 5307 ^{12}	1.715 8845
19	2.446 6267	4.467 7531	8.004 7243	1.408 8315	2.438 3622
20	2.691 2894	5.137 9161	9.605 6692	1.761 0394	3.169 8708
21	2.691 2894	5.382 5787	1.052 0495 ^{12}	2.012 6164	3.773 6558
22			1.052 0495	2.104 0990	4.116 7154
23					4.116 7154

n	$m = 46$	$m = 47$	$m = 48$	$m = 49$	$m = 50$
1	4.6 ^{1}	4.7 ^{1}	4.8 ^{1}	4.9 ^{1}	5.0 ^{1}
2	1.035 ^{3}	1.081 ^{3}	1.128 ^{3}	1.176 ^{3}	1.225 ^{3}
3	1.518 0 ^{4}	1.621 5 ^{4}	1.729 6 ^{4}	1.842 4 ^{4}	1.960 0 ^{4}
4	1.631 85 ^{5}	1.783 65 ^{5}	1.945 80 ^{5}	2.118 76 ^{5}	2.303 00 ^{5}
5	1.370 754 ^{6}	1.533 939 ^{6}	1.712 304 ^{6}	1.906 884 ^{6}	2.118 760 ^{6}
6	9.366 819	1.073 7573 ^{7}	1.227 1512 ^{7}	1.398 3816 ^{7}	1.589 0700 ^{7}
7	5.352 4680 ^{7}	6.289 1499	7.362 9072	8 590 0584	9 988 4400
8	2.609 3282 ^{8}	3.144 5750 ^{8}	3.773 4899 ^{8}	4 509 7807 ^{8}	5 368 7865 ^{8}
9	1.101 7163 ^{9}	1.362 6491 ^{9}	1.677 1066 ^{9}	2.054 4556 ^{9}	2.505 4337 ^{9}
10	4.076 3504	5.178 0668	6.540 7159	8.217 8225	1.027 2278 ^{10}
11	1.334 0783 ^{10}	1.741 7134 ^{10}	2.259 5200 ^{10}	2.913 5916 ^{10}	3.735 3739
12	3.891 0618	5.225 1401	6.966 8534	9.226 3735	1.213 9965 ^{11}
13	1.017 6623 ^{11}	1.406 7685 ^{11}	1.929 2825 ^{11}	2.625 9678 ^{11}	3.548 6052
14	2.398 7754	3.416 4377	4.823 2062	6.752 4887	9.378 4566
15	5.117 3876	7.516 1630	1.093 2601 ^{12}	1.575 5807 ^{12}	2.250 8296 ^{12}
16	9.914 9385	1.503 2326 ^{12}	2.254 8489	3.348 1090	4.923 6897
17	1.749 6950 ^{12}	2.741 1889	4.244 4215	6.499 2704	9.847 3794
18	2.818 9531	4.568 6481	7.309 8370	1.155 4258 ^{13}	1.805 3529 ^{13}
19	4.154 2467	6.973 1998	1.154 1848 ^{13}	1.885 1685	3.040 5943
20	5.608 2330	9.762 4797	1.673 5679	2.827 7527	4.712 9212
21	6.943 5266	1.255 1760 ^{13}	2.231 4239	3.904 9919	6.732 7446
22	7.890 3711	1.483 3898	2.738 5657	4.969 9897	8.874 9815
23	8.233 4307	1.612 3802	3.095 7700	5.834 3357	1.080 4325 ^{14}
24		1.612 3802	3.224 7604	6.320 5303	1.215 4866
25				6.320 5303	1.264 1061

n	$m = 51$	$m = 52$	$m = 53$	$m = 54$	$m = 55$
1	5.1^{1}	5.2^{1}	5.3^{1}	5.4^{1}	5.5^{1}
2	1.275^{3}	1.326^{3}	1.378^{3}	1.431^{3}	1.485^{3}
3	$2.082\ 5^{4}$	$2.210\ 0^{4}$	$2.342\ 6^{4}$	$2.480\ 4^{4}$	$2.623\ 5^{4}$
4	$2.499\ 00^{5}$	$2.707\ 25^{5}$	$2.928\ 25^{5}$	$3.162\ 51^{5}$	$3.410\ 55^{5}$
5	$2.349\ 060^{6}$	$2.598\ 960^{6}$	$2.869\ 685^{6}$	$3.162\ 510^{6}$	$3.478\ 761^{6}$
6	$1.800\ 9460^{7}$	$2.035\ 8520^{7}$	$2.295\ 7480^{7}$	$2.582\ 7165^{7}$	$2.898\ 9675^{7}$
7	$1.157\ 7510^{8}$	$1.337\ 8456^{8}$	$1.541\ 4308^{8}$	$1.771\ 0056^{8}$	$2.029\ 2773^{8}$
8	$6.367\ 6305$	$7.525\ 3815$	$8.863\ 2271$	$1.040\ 4658^{9}$	$1.217\ 5664^{9}$
9	$3.042\ 3124^{9}$	$3.679\ 0754^{9}$	$4.431\ 6136^{9}$	$5.317\ 9363$	$6.358\ 4021$
10	$1.277\ 7712^{10}$	$1.582\ 0024^{10}$	$1.949\ 9100^{10}$	$2.393\ 0713^{10}$	$2.924\ 8649^{10}$
11	$4.762\ 6017$	$6.040\ 3729$	$7.622\ 3753$	$9.572\ 2853$	$1.196\ 5357^{11}$
12	$1.587\ 5339^{11}$	$2.063\ 7941^{11}$	$2.667\ 8314^{11}$	$3.430\ 0689^{11}$	$4.387\ 2974$
13	$4.762\ 6017$	$6.350\ 1356$	$8.413\ 9297$	$1.108\ 1761^{12}$	$1.451\ 1830^{12}$
14	$1.292\ 7062^{12}$	$1.768\ 9663^{12}$	$2.403\ 9799^{12}$	$3.245\ 3729$	$4.353\ 5490$
15	$3.188\ 6752$	$4.481\ 3814$	$6.250\ 3478$	$8.654\ 3277$	$1.189\ 9701^{13}$
16	$7.174\ 5193$	$1.036\ 3195^{13}$	$1.484\ 4576^{13}$	$2.109\ 4924^{13}$	$2.974\ 9251$
17	$1.477\ 1069^{13}$	$2.194\ 5588$	$3.230\ 8783$	$4.715\ 3359$	$6.824\ 8282$
18	$2.790\ 0908$	$4.267\ 1977$	$6.461\ 7566$	$9.692\ 6349$	$1.440\ 7971^{14}$
19	$4.845\ 9472$	$7.636\ 0381$	$1.190\ 3236^{14}$	$1.836\ 4992^{14}$	$2.805\ 7627$
20	$7.753\ 5156$	$1.259\ 9463^{14}$	$2.023\ 5501$	$3.213\ 8737$	$5.050\ 3729$
21	$1.144\ 5666^{14}$	$1.919\ 9181$	$3.179\ 8644$	$5.203\ 4145$	$8.417\ 2882$
22	$1.560\ 7726$	$2.705\ 3392$	$4.625\ 2573$	$7.805\ 1218$	$1.300\ 8536^{15}$
23	$1.967\ 9307$	$3.528\ 7033$	$6.234\ 0425$	$1.085\ 9300^{15}$	$1.866\ 4422$
24	$2.295\ 9191$	$4.263\ 8498$	$7.792\ 5531$	$1.402\ 6596$	$2.488\ 5895$
25	$2.479\ 5927$	$4.775\ 5118$	$9.039\ 3616$	$1.683\ 1915$	$3.085\ 8510$
26	$2.479\ 5927$	$4.959\ 1853$	$9.734\ 6971$	$1.877\ 4059$	$3.560\ 5973$
27			$9.734\ 6971$	$1.946\ 9394$	$3.824\ 3453$
28					$3.824\ 3453$

n	$m = 56$	$m = 57$	$m = 58$	$m = 59$	$m = 60$
1	5.6^{1}	5.7^{1}	5.8^{1}	5.9^{1}	6.0^{1}
2	1.540^{3}	1.596^{3}	1.653^{3}	1.711^{3}	1.770^{3}
3	$2.772\ 0^{4}$	$2.926\ 0^{4}$	$3.085\ 6^{4}$	$3.250\ 9^{4}$	$3.422\ 0^{4}$
4	$3.672\ 90^{5}$	$3.950\ 10^{5}$	$4.242\ 70^{5}$	$4.551\ 26^{5}$	$4.876\ 35^{5}$
5	$3.819\ 816^{6}$	$4.187\ 106^{6}$	$4.582\ 116^{6}$	$5.006\ 386^{6}$	$5.461\ 512^{6}$
6	$3.246\ 8436^{7}$	$3.628\ 8252^{7}$	$4.047\ 5358^{7}$	$4.505\ 7474^{7}$	$5.006\ 3860^{7}$
7	$2.319\ 1740^{8}$	$2.643\ 8584^{8}$	$3.006\ 7409^{8}$	$3.411\ 4945^{8}$	$3.862\ 0692^{8}$
8	$1.420\ 4941^{9}$	$1.652\ 4115^{9}$	$1.916\ 7973^{9}$	$2.217\ 4714^{9}$	$2.558\ 6208^{9}$
9	$7.575\ 9684$	$8.996\ 4625$	$1.064\ 8874^{10}$	$1.256\ 5671^{10}$	$1.478\ 3143^{10}$
10	$3.560\ 7051^{10}$	$4.318\ 3020^{10}$	$5.217\ 9482$	$6.282\ 8356$	$7.539\ 4028$

n	$m = 56$	$m = 57$	$m = 58$	$m = 59$	$m = 60$
11	1.489 0222 ^{11}	1.845 0927 ^{11}	2.276 9229 ^{11}	2.798 7177 ^{11}	3.427 0013 ^{11}
12	5.583 8331	7.072 8552	8.917 9479	1.119 4871 ^{12}	1.399 3588 ^{12}
13	1.889 9127 ^{12}	2.448 2960 ^{12}	3.155 5816 ^{12}	4.047 3764	5.166 8634
14	5.804 7320	7.694 6447	1.014 2941 ^{13}	1.329 8522 ^{13}	1.734 5899 ^{13}
15	1.625 3249 ^{13}	2.205 7981 ^{13}	2.975 2626	3.989 5567	5.319 4089
16	4.164 8952	5.790 2201	7.996 0183	1.097 1281 ^{14}	1.496 0838 ^{14}
17	9.799 7534	1.396 4649 ^{14}	1.975 4869 ^{14}	2.775 0887	3.872 2168
18	2.123 2799 ^{14}	3.103 2552	4.499 7201	6.475 2070	9.250 2957
19	4.246 5598	6.369 8397	9.473 0949	1.397 2815 ^{15}	2.044 8022 ^{15}
20	7.856 1356	1.210 2695 ^{15}	1.847 2535 ^{15}	2.794 5630	4.191 8445
21	1.346 7661 ^{15}	2.132 3797	3.342 6492	5.189 9027	7.984 4657
22	2.142 5824	3.489 3485	5.621 7282	8.964 3774	1.415 4280 ^{16}
23	3.167 2958	5.309 8782	8.799 2268	1.442 0955 ^{16}	2.338 5332
24	4.355 0317	7.522 3275	1.283 2206 ^{16}	2.163 1432	3.605 2387
25	5.574 4406	9.929 4723	1.745 1800	3.028 4005	5.191 5438
26	6.646 4484	1.222 0889 ^{16}	2.215 0361	3.960 2161	6.988 6167
27	7.384 9427	1.403 1391	2.625 2280	4.840 2641	8.800 4802
28	7.648 6906	1.503 3633	2.906 5024	5.531 7304	1.037 1995 ^{17}
29		1.503 3633	3.006 7266	5.913 2291	1.149 4960
30				5.913 2291	1.182 6458

n	$m = 61$	$m = 62$	$m = 63$	$m = 64$	$m = 65$
1	6.1 ^{1}	6.2 ^{1}	6.3 ^{1}	6.4 ^{1}	6.5 ^{1}
2	1.830 ^{3}	1.891 ^{3}	1.953 ^{3}	2.016 ^{3}	2.080 ^{3}
3	3.599 0 ^{4}	3.782 0 ^{4}	3.971 1 ^{4}	4.166 4 ^{4}	4.368 0 ^{4}
4	5.218 55 ^{5}	5.578 45 ^{5}	5.956 65 ^{5}	6.353 76 ^{5}	6.770 40 ^{5}
5	5.949 147 ^{6}	6.471 002 ^{6}	7.028 847 ^{6}	7.624 512 ^{6}	8.259 888 ^{6}
6	5.552 5372 ^{7}	6.147 4519 ^{7}	6.794 5521 ^{7}	7.497 4368 ^{7}	8.259 8880 ^{7}
7	4.362 7078 ^{8}	4.917 9615 ^{8}	5.532 7067 ^{8}	6.212 1619 ^{8}	6.961 9056 ^{8}
8	2.944 8278 ^{9}	3.381 0985 ^{9}	3.872 8947 ^{9}	4.426 1654 ^{9}	5.047 3816 ^{9}
9	1.734 1764 ^{10}	2.028 6591 ^{10}	2.366 7690 ^{10}	2.754 0585 ^{10}	3.196 6750 ^{10}
10	9.017 7170	1.075 1893 ^{11}	1.278 0553 ^{11}	1.514 7321 ^{11}	1.790 1380 ^{11}
11	4.180 9415 ^{11}	5.082 7132	6.157 9026	7.435 9578	8.950 6900
12	1.742 0590 ^{12}	2.160 1531 ^{12}	2.668 4244 ^{12}	3.284 2147 ^{12}	4.027 8105 ^{12}
13	6.566 2223	8.308 2812	1.046 8434 ^{13}	1.313 6859 ^{13}	1.642 1074 ^{13}
14	2.251 2762 ^{13}	2.907 8984 ^{13}	3.738 7266	4.785 5700	6.099 2559
15	7.053 9988	9.305 2750	1.221 3173 ^{14}	1.595 1900 ^{14}	2.073 7470 ^{14}
16	2.028 0247 ^{14}	2.733 4245 ^{14}	3.663 9520	4.885 2694	6.480 4594
17	5.368 3005	7.396 3252	1.012 9750 ^{15}	1.379 3702 ^{15}	1.867 8971 ^{15}
18	1.312 2512 ^{15}	1.849 0813 ^{15}	2.588 7138	3.601 6888	4.981 0590
19	2.969 8318	4.282 0830	6.131 1643	8.719 8781	1.232 1567 ^{16}
20	6.236 6467	9.206 4785	1.348 8561 ^{16}	1.961 9726 ^{16}	2.833 9604

n	$m = 61$	$m = 62$	$m = 63$	$m = 64$	$m = 65$
21	1 217 6310 16	1.841 2957 16	2 761 9435 16	4.110 7997 16	6 072 7723 16
22	2.213 8746	3.431 5056	5.272 8013	8.034 7448	1.214 5545 17
23	3.753 9613	5.967 8358	9.399 3415	1.467 2143 17	2.270 6888
24	5.943 7720	9.697 7332	1.566 5569 17	2.506 4911	3.973 7053
25	8.796 7825	1.474 0555 17	2.443 8288	4.010 3857	6.516 8767
26	1.218 0160 17	2.097 6943	3.571 7498	6.015 5785	1.002 5964 18
27	1 578 9297	2 796 9257	4.894 6200	8.466 3698	1.448 1948
28	1.917 2475	3.496 1572	6.293 0829	1.118 7703 18	1.965 4073
29	2.181 6954	4.098 9429	7.595 1000	1.388 8183	2.507 5886
30	2.327 1418	4.508 8372	8.607 7801	1.620 2880	3.009 1063
31	2 327 1418	4.654 2835	9.163 1207	1.777 0901	3.397 3781
32			9.163 1207	1.832 6241	3.609 7142
33					3.609 7142

n	$m = 66$	$m = 67$	$m = 68$	$m = 69$	$m = 70$
1	6.6^{1}	6.7^{1}	6.8^{1}	6.9^{1}	7.0^{1}
2	2.145^{3}	2.211^{3}	2.278^{3}	2.346^{3}	2.415^{3}
3	$4.576\ 0^{4}$	$4.790\ 5^{4}$	$5.011\ 6^{4}$	$5.239\ 4^{4}$	$5.474\ 0^{4}$
4	$7.207\ 20^{5}$	$7.664\ 80^{5}$	$8.143\ 85^{5}$	$8.645\ 01^{5}$	$9.168\ 95^{5}$
5	$8.936\ 928^{6}$	$9.657\ 648^{6}$	$1.042\ 4128^{7}$	$1.123\ 8513^{7}$	$1.210\ 3014^{7}$
6	9 085 8768 7	9 979 5696 7	1.094 5334 8	1.198 7747 8	1.311 1599 8
7	7.787 8944 8	8.696 4821 8	9.694 4390	1.078 8972 9	1.198 7747 9
8	5 743 5721 9	6 522 3616 9	7 392 0098 9	8 361 4537	9 449 3599
9	3 701 4131 10	4 275 7704 10	4 928 0065 10	5 667 2075 10	6 503 3529 10
10	2.109 8055 11	2 479 9468 11	2 907 5238 11	3.400 3245 11	3 967 0452 11
11	1.074 0828 12	1.285 0633 12	1.533 0580 12	1.823 8104 12	2.163 8429 12
12	4 922 8795	5 996 9623	7 282 0256	8 815 0836	1 063 8894 13
13	2.044 8884 13	2 537 1763 13	3 136 8726 13	3 865 0751 13	4 746 5835
14	7.741 3632	9.786 2516	1.232 3428 14	1.546 0301 14	1.932 5376 14
15	2 683 6726 14	3.457 8089 14	4.436 4341	5 668 7769	7.214 8069
16	8.554 2064	1.123 7879 15	1.469 5688 15	1.913 2122 15	2 480 0899 15
17	2 515 9430 15	3.371 3637	4.495 1516	5.964 7204	7 877 9326
18	6.848 9561	9.364 8991	1.273 6263 16	1.723 1414 16	2.319 6135 16
19	1.730 2626 16	2 415 1582 16	3 351 6481	4 625 2744	6 348 4158
20	4 066 1171	5.796 3797	8 211 5379	1.156 3186 17	1 618 8460 17
21	8 906 7327	1.297 2850 17	1.876 9229 17	2.698 0767	3 854 3953
22	1.821 8317 17	2.712 5049	4.009 7899	5.886 7129	8.584 7896
23	3 485 2432	5.307 0749	8.019 5798	1.202 9370 18	1.791 6083 18
24	6 244 3941	9.729 6373	1.503 6712 18	2.305 6292	3.508 5662
25	1.049 0582 18	1.673 4976 18	2.646 4613	4.150 1326	6.455 7618

n	$m = 66$	$m = 67$	$m = 68$	$m = 69$	$m = 70$
26	1.654 2841 18	2.703 3234 18	4.376 8399 18	7.023 3013 18	1.117 3434 19
27	2.450 7913	4.105 0753	6.808 4177	1.118 5258 19	1.820 8559
28	3.413 6021	5.864 3934	9.969 4687	1.677 7886	2.796 3174
29	4.472 9959	7.886 5980	1.375 0991 19	2.372 0460	4.049 8346
30	5.516 6949	9.989 6908	1.787 6289	3.162 7280	5.534 7740
31	6.406 4844	1.192 3179 19	2.191 2870	3.978 9159	7.141 6438
32	7.007 0923	1.341 3577	2.533 6756	4.724 9626	8.703 8785
33	7.219 4284	1.422 6521	2.764 0097	5.297 6853	1.002 2648 20
34		1.422 6521	2.845 3041	5.609 3139	1.040 6999
35				5.609 3139	1.121 8628

n	$m = 71$	$m = 72$	$m = 73$	$m = 74$	$m = 75$
1	7.1 1	7.2 1	7.3 1	7.4 1	7.5 1
2	2.485 3	2.556 3	2.628 3	2.701 3	2.775 3
3	5.715 5 4	5.964 0 4	6.219 6 4	6.482 4 4	6.752 5 4
4	9.716 35 5	1.028 790 6	1.088 430 6	1.150 626 6	1.215 450 6
5	1.301 9909 7	1.399 1544 7	1.502 0334 7	1.610 8764 7	1.725 9390 7
6	1.432 1900 8	1.562 3891 8	1.702 3045 8	1.852 5079 8	2.013 5955 8
7	1.329 8907 9	1.473 1097 9	1.629 3486 9	1.799 5791 9	1.984 8299 9
8	1.063 9126 10	1.196 9016 10	1.344 2126 10	1.507 1475 10	1.687 1054 10
9	7.447 3879	8.511 3005	9.708 2021	1.105 2415 11	1.255 9562 11
10	4.617 3805 11	5.362 1193 11	6.213 2494 11	7.184 0696	8.289 3111
11	2.560 5474 12	3.022 2854 12	3.558 4974 12	4.179 8223 12	4.898 2293 12
12	1.280 2737 13	1.536 3284 13	1.838 5570 13	2.194 4067 13	2.612 3889 13
13	5.810 4729	7.090 7466	8.627 0750	1.046 5632 14	1.266 0039 14
14	2.407 1959 14	2.988 2432 14	3.697 3179 14	4.560 0254	5.606 5886
15	9.147 3445	1.155 4540 15	1.454 2784 15	1.824 0101 15	2.280 0127 15
16	3.201 5706 15	4.116 3050	5.271 7591	6.726 0374	8.550 0476
17	1.035 8022 16	1.355 9593 16	1.767 5898 16	2.294 7657 16	2.967 3695 16
18	3.107 4067	4.143 2090	5.499 1683	7.266 7581	9.561 5238
19	8.668 0293	1.177 5436 17	1.591 8645 17	2.141 7813 17	2.868 4571 17
20	2.253 6876 17	3.120 4906	4.298 0342	5.889 8987	8.031 6800
21	5.473 2414	7.726 9290	1.084 7420 18	1.514 5454 18	2.103 5352 18
22	1.243 9185 18	1.791 2426 18	2.563 9355	3.648 6775	5.163 2228
23	2.650 0872	3.894 0057	5.685 2483	8.249 1839	1.189 7861 19
24	5.300 1744	7.950 2617	1.184 4267 19	1.752 9516 19	2.577 8700
25	9.964 3279	1.526 4502 19	2.321 4764	3.505 9031	5.258 8547
26	1.762 9196 19	2.759 3524	4.285 8026	6.607 2790	1.011 3182 20
27	2.938 1993	4.701 1188	7.460 4712	1.174 6274 20	1.835 3553
28	4.617 1703	7.555 3695	1.225 6488 20	1.971 6960	3.146 3233
29	6.846 1490	1.146 3319 20	1.901 8689	3.127 5177	5.099 2137
30	9.584 6086	1.643 0758	2.789 4077	4.691 2766	7.818 7943

n	$m = 71$	$m = 72$	$m = 73$	$m = 74$	$m = 75$
31	1 267 6418 20	2.226 1027 20	3.869 1784 20	6.658 5861 20	1.134 9863 21
32	1 584 5522	2.852 1940	5.078 2967	8.947 4751	1.560 6061
33	1 872 6526	3.457 2049	6.309 3989	1.138 7696 21	2.033 5171
34	2.092 9647	3.965 6174	7.422 8222	1.373 2221	2.511 9917
35	2 212 5627	4.305 5274	8.271 1448	1.569 3967	2.942 6188
36	2 212 5627	4.425 1254	8.730 6528	1.700 1798	3.269 5765
37			8.730 6528	1.746 1306	3.446 3103
38					3.446 3103

n	$m = 76$	$m = 77$	$m = 78$	$m = 79$	$m = 80$
1	7.6 1	7.7 1	7.8 1	7.9 1	8.0 1
2	2.850 3	2.926 3	3.003 3	3.081 3	3.160 3
3	7.030 0 4	7.315 0 4	7.607 6 4	7.907 9 4	8.216 0 4
4	1.282 975 6	1.353 275 6	1.426 425 6	1.502 501 6	1.581 580 6
5	1.847 4840 7	1.975 815 7	2.111 1090 7	2.253 7515 7	2.404 0016 7
6	2.186 1894 8	2.370 9378 8	2.568 5160 8	2.779 6269 8	3.005 0020 8
7	2.186 1894 9	2.404 8083 9	2.641 9021 9	2.898 7537 9	3.176 7164 9
8	1.885 5884 10	2.104 2073 10	2.344 6881 10	2.608 8783 10	2.898 7537 10
9	1.424 6668 11	1.613 2256 11	1.823 6463 11	2.058 1151 11	2.319 0030 11
10	9.545 2673	1.096 9934 12	1.258 3160 12	1.440 6806 12	1.646 4921 12
11	5.727 1604 12	6.681 6871	7.778 6805	9.036 9965	1.047 7677 13
12	3.102 2119 13	3.674 9279 13	4.343 0966 13	5.120 9647 13	6.024 6643
13	1.527 2428 14	1.837 4640 14	2.204 9567 14	2.639 2664 14	3.151 3629 14
14	6.872 5924	8.399 8352	1.023 7299 15	1.244 2256 15	1.508 1522 15
15	2.840 6715 15	3.527 9308 15	4.367 9143	5.392 5442	6.635 8698
16	1.083 0060 16	1.367 0732 16	1.719 8663 16	2.156 6577 16	2.695 8221 16
17	3.822 3742	4.905 3802	6.272 4534	7.992 3197	1.014 8977 17
18	1.252 8893 17	1.635 1267 17	2.125 6648 17	2.752 9101 17	3.552 1421
19	3.824 6095	5.077 4988	6.712 6256	8.838 2904	1.159 1200 18
20	1.090 0137 18	1.472 4747 18	1.980 2245 18	2.651 4871 18	3.535 3161
21	2.906 7032	3.996 7169	5.469 1916	7.449 4162	1.010 0903 19
22	7.266 7581	1.017 3461 19	1.417 0178 19	1.963 9370 19	2.708 8786
23	1.706 1084 19	2.432 7842	3.450 1304	4.867 1482	6.831 0852
24	3.767 6561	5 473 7645	7.906 5487	1.135 6679 20	1.622 3827 20
25	7.836 7247	1.160 4381 20	1.707 8145 20	2.498 4694	3.634 1373
26	1.537 2037 20	2.320 8762	3.481 3142	5.189 1288	7.687 5982
27	2.846 6735	4.383 8772	6.704 7533	1.018 6068 21	1.537 5196 21
28	4.981 6786	7.828 3521	1.221 2229 21	1.891 6983	2.910 3050
29	8.245 5370	1.322 7216 21	2.105 5568	3.326 7797	5.218 4780
30	1.291 8008 21	2.116 3545	3.439 0761	5.544 6328	8.871 4125

n	$m = 76$	$m = 77$	$m = 78$	$m = 79$	$m = 80$
31	1.916 8657 21	3.208 6665 21	5.325 0210 21	8.764 0971 21	1.430 8730 22
32	2.695 5924	4.612 4581	7.821 1246	1.314 6146 22	2.191 0243
33	3.594 1232	6.289 7156	1.090 2174 22	1.872 3298	3.186 9444
34	4.545 5087	8.139 6319	1.442 9348	2.533 1521	4.405 4819
35	5.454 6105	1.000 0119 22	1.813 9751	3.256 9099	5.790 0620
36	6.212 1953	1.166 6806	2.166 6925	3.980 6676	7.237 5775
37	6.715 8868	1.292 8082	2.459 4888	4.626 1813	8.606 8489
38	6.892 6206	1.360 8507	2.653 6589	5.113 1477	9.739 3290
39		1.360 8507	2.721 7015	5.375 3604	1.048 8508 23
40				5.375 3604	1.075 0721

n	$m = 81$	$m = 82$	$m = 83$	$m = 84$	$m = 85$
1	8.1 1	8.2 1	8.3 1	8.4 1	8.5 1
2	3.240 3	3.321 3	3.403 3	3.486 3	3.570 3
3	8.532 0 4	8.856 0 4	9.188 1 4	9.528 4 4	9.877 0 4
4	1.663 740 6	1.749 060 6	1.837 620 6	1.929 501 6	2.024 785 6
5	2.562 1596 7	2.728 5336 7	2.903 4396 7	3.087 2016 7	3.280 1517 7
6	3.245 4022 8	3.501 6181 8	3.774 4715 8	4 064 8154 8	4 373 5356 8
7	3.477 2166 9	3.801 7568 9	4 151 9186 9	4 529 3658 9	4 935 8473 9
8	3.216 4254 10	3.564 1470 10	3.944 3227 10	4 359 5146 10	4 812 4511 10
9	2 608 8783 11	2.930 5209 11	3 286 9356 11	3 681 3679 11	4 117 3193 11
10	1.878 3924 12	2.139 2802 12	2.432 3323 12	2.761 0259 12	3.129 1627 12
11	1.212 4169 13	1.400 2562 13	1.614 1842 13	1.857 4174 13	2.133 5200 13
12	7.072 4320	8 284 8489	9 685 1051	1.129 9289 14	1 315 6707 14
13	3.753 8293 14	4 461 0725 14	5.289 5574 14	6 258 0679	7 387 9968
14	1.823 2885 15	2.198 6714 15	2.644 7787 15	3.173 7344 15	3.799 5412 15
15	8.144 0220	9.967 3106	1.216 5982 16	1.481 0761 16	1.798 4495 16
16	3.359 4091 16	4.173 8113 16	5.170 5424	6.387 1406	7.868 2166
17	1.284 4799 17	1.620 4206 17	2.037 8020 17	2.554 8562 17	3.193 5703 17
18	4.567 0398	5.851 5198	7.471 9406	9.509 7426	1.206 4599 18
19	1.514 3343 18	1 971 0382 18	2.556 1902 18	3.303 3843 18	4 254 3585
20	4.694 4362	6.208 7704	8.179 8087	1.073 5999 19	1.403 9383 19
21	1.363 6219 19	1.833 0656 19	2.453 9426 19	3.271 9235	4.345 5234
22	3.718 9689	5.082 5909	6.915 6564	9.369 5990	1.264 1523 20
23	9.539 9638	1.325 8933 20	1.834 1524 20	2.525 7180 20	3.462 6779
24	2.305 4912 20	3.259 4876	4.585 3809	6.419 5333	8.945 2513
25	5.256 5200	7.562 0113	1.082 1499 21	1.540 6880 21	2.182 6413 21
26	1.132 1735 21	1.657 8256 21	2.414 0267	3.496 1766	5.036 8646
27	2.306 2794	3.438 4530	5.096 2785	7.510 3052	1.100 6482 22
28	4.447 8246	6.754 1041	1.019 2557 22	1.528 8836 22	2.279 9141
29	8.128 7830	1.257 6608 22	1.933 0712	2.952 3269	4.481 2104
30	1.408 9890 22	2.221 8673	3.479 5281	5.412 5993	8.364 9262

n	$m = 81$	$m = 82$	$m = 83$	$m = 84$	$m = 85$
31	2.318 0142 22	3.727 0032 22	5.948 8706 22	9.428 3988 22	1.484 0998 23
32	3.621 8973	5.939 9115	9.666 9148	1.561 5785 23	2.504 4184
33	5.377 9687	8.999 8659	1.493 9777 23	2.460 6692	4.022 2478
34	7.592 4263	1.297 0395 23	2.197 0261	3.691 0038	6.151 6730
35	1.019 5544 23	1.778 7970	3.075 8365	5.272 8626	8.963 8664
36	1.302 7639	2.322 3183	4.101 1154	7.176 9519	1.244 9815 24
37	1.584 4426	2.887 2066	5.209 5249	9.310 6403	1.648 7592
38	1.834 6178	3.419 0604	6.306 2670	1.151 5792 24	2.082 6432
39	2.022 7837	3.857 4015	7.276 4619	1.358 2729	2.509 8520
40	2.123 9229	4.146 7066	8.004 1081	1.528 0570	2.886 3299
41	2.123 9229	4.247 8458	8.394 5524	1.639 8661	3.167 9231
42			8.394 5524	1.678 9105	3.318 7765
43					3 318 7765

n	$m = 86$	$m = 87$	$m = 88$	$m = 89$	$m = 90$
1	8 6 1	8 7 1	8 8 1	8 9 1	9 0 1
2	3.655 3	3.741 3	3.828 3	3.916 3	4.005 3
3	1 023 40 5	1 059 95 5	1.097 36 5	1.135 64 5	1.174 80 5
4	2 123 555 6	2 225 895 6	2 331 890 6	2.441 626 6	2.555 190 6
5	3 482 6302 7	3 694 9857 7	3.917 5752 7	4.150 7642 7	4.394 9268 7
6	4 701 5508 8	5 049 8138 8	5.419 3124 8	5.811 0699 8	6.226 1463 8
7	5 373 2009 9	5 843 3560 9	6.348 3373 9	6.890 2686 9	7 471 3756 9
8	5 306 0359 10	5.843 3560 10	6.427 6916 10	7.062 5253 10	7.751 5521 10
9	4 598 5644 11	5.129 1680 11	5.713 5036 11	6.356 2728 11	7.062 5253 11
10	3 540 8946 12	4.000 7510 12	4.513 6678 12	5.085 0182 12	5 720 6455 12
11	2.446 4363 13	2.800 5257 13	3.200 6008 13	3.651 9676 13	4.160 4694 13
12	1 529 0227 14	1 773 6663 14	2.053 7189 14	2.373 7790 14	2.738 9757 14
13	8 703 6675	1.023 2690 15	1.200 6356 15	1.406 0075 15	1.643 3854 15
14	4 538 3499 15	5 408 7077	6.431 9767	7 632 6123	9.038 6199
15	2.178 4036 16	2.632 2377 16	3.173 1085 16	3.816 3062 16	4.579 5674 16
16	9 666 6661	1.184 5070 17	1.447 7308 17	1.765 0416 17	2.146 6722 17
17	3.980 3919 17	4.947 0586	6.131 5655	7.579 2963	9.344 3379
18	1.525 8169 18	1.923 8561 18	2.418 5620 18	3.031 7185 18	3.789 6481 18
19	5.460 8184	6.986 6353	8.910 4914	1.132 9053 19	1.436 0772 19
20	1.829 3742 19	2.375 4560 19	3.074 1195 19	3.965 1687	5.098 0740
21	5.749 4617	7.578 8358	9.954 2919	1.302 8411 20	1.699 3580 20
22	1.698 7046 20	2.273 6508 20	3 031 5343 20	4.026 9635	5.329 8047
23	4.726 8302	6.425 5347	8.699 1855	1.173 0720 21	1.575 7683 21
24	1.240 7929 21	1.713 4759 21	2.356 0294 21	3.225 9480	4.399 0199
25	3.077 1664	4.317 9593	6.031 4353	8.387 4647	1.161 3413 22

n	$m = 86$	$m = 87$	$m = 88$	$m = 89$	$m = 90$
26	$7.219\ 5059^{21}$	$1.029\ 6672^{22}$	$1.461\ 4632^{22}$	$2.064\ 6067^{22}$	$2.903\ 3532^{22}$
27	$1.604\ 3346^{22}$	2.326 2852	3.355 9524	4 817 4156	6 882 0223
28	3.380 5623	4.984 8969	$7.311\ 1821$	$1.066\ 7135^{23}$	$1.548\ 4550^{23}$
29	6.761 1245	$1.014\ 1687^{23}$	$1.512\ 6584^{23}$	2.243 7766	3.310 4900
30	$1.284\ 6137^{23}$	1.960 7261	2.974 8948	4.487 5532	6.731 3297
31	2.320 5924	3.605 2061	5.565 9322	8.540 8270	$1.302\ 8380^{24}$
32	3.988 5182	6.309 1106	9.914 3167	$1.548\ 0249^{24}$	2.402 1076
33	6.526 6662	$1.051\ 5184^{24}$	$1.682\ 4295^{24}$	2.673 8612	4.221 8861
34	$1.017\ 3921^{24}$	1.670 0587	2.721 5771	4.404 0066	7.077 8678
35	1.511 5539	2.528 9460	4.199 0047	6.920 5819	$1.132\ 4589^{25}$
36	2.141 3681	3.652 9220	6.181 8681	$1.038\ 0873^{25}$	1.730 1455
37	2.893 7407	5.035 1088	8.688 0308	1.486 9899	2.525 0772
38	3.731 4024	6.625 1431	$1.166\ 0252^{25}$	2.034 8283	3.521 8182
39	4.592 4952	8.323 8977	1.494 9041	2.660 9293	4.695 7575
40	5.396 1819	9.988 6771	1.831 2575	3.326 1616	5.987 0908
41	6.054 2530	$1.145\ 0435^{25}$	2.143 9112	3.975 1687	7.301 3302
42	6.486 6996	1.254 0953	2.399 1387	4.543 0499	8.518 2186
43	6.637 5531	1.312 4253	2.566 5205	4.965 6593	9.508 7092
44		1.312 4253	2.624 8505	5.191 3711	$1.015\ 7030^{26}$
45				5.191 3711	1.038 2742

n	$m = 91$	$m = 92$	$m = 93$	$m = 94$	$m = 95$
1	9.1^{1}	9.2^{1}	9.3^{1}	9.4^{1}	9.5^{1}
2	4.095^{3}	4.186^{3}	4.278^{3}	4.371^{3}	4.465^{3}
3	$1.214\ 85^{5}$	$1.255\ 80^{5}$	$1.297\ 66^{5}$	$1.340\ 44^{5}$	$1.384\ 15^{5}$
4	$2.672\ 670^{6}$	$2.794\ 155^{6}$	$2.919\ 735^{6}$	$3.049\ 501^{6}$	$3.183\ 545^{6}$
5	$4.650\ 4458^{7}$	$4.917\ 7128^{7}$	$5.197\ 1283^{7}$	$5.489\ 1018^{7}$	$5.794\ 0519^{7}$
6	$6.665\ 6390^{8}$	$7.130\ 6836^{8}$	$7.622\ 4548^{8}$	$8.142\ 1677^{8}$	$8.691\ 0779^{8}$
7	$8.093\ 9902^{9}$	$8.760\ 5841^{9}$	$9.473\ 6224^{9}$	$1.023\ 5868^{10}$	$1.105\ 0085^{10}$
8	$8.498\ 6897^{10}$	$9.308\ 0887^{10}$	$1.018\ 4144^{11}$	$1.113\ 1506^{11}$	$1.215\ 5093^{11}$
9	$7.837\ 6805^{11}$	$8.687\ 5495^{11}$	$9.618\ 3583$	$1.063\ 6773^{12}$	$1.174\ 9923^{12}$
10	$6.426\ 8980^{12}$	$7.210\ 6661^{12}$	$8.079\ 4210^{12}$	9.041 2569	$1.010\ 4934^{13}$
11	$4.732\ 5340^{13}$	$5.375\ 2238^{13}$	$6.096\ 2904^{13}$	$6.904\ 2325^{13}$	7.808 3582
12	$3.155\ 0227^{14}$	$3.628\ 2761^{14}$	$4.165\ 7984^{14}$	$4.775\ 4275^{14}$	$5.465\ 8507^{14}$
13	$1.917\ 2830^{15}$	$2.232\ 7853^{15}$	$2.595\ 6129^{15}$	$3.012\ 1927^{15}$	$3.489\ 7355^{15}$
14	$1.068\ 2005^{16}$	$1.259\ 9288^{16}$	$1.483\ 2074^{16}$	$1.742\ 7686^{16}$	$2.043\ 9879^{16}$
15	5.483 4294	6.551 6299	7.811 5587	9.294 7661	$1.103\ 7535^{17}$
16	$2.604\ 6290^{17}$	$3.152\ 9719^{17}$	$3.808\ 1349^{17}$	$4.589\ 2908^{17}$	5.518 7674
17	$1.149\ 1010^{18}$	$1.409\ 5639^{18}$	$1.724\ 8611^{18}$	$2.105\ 6746^{18}$	$2.564\ 6037^{18}$
18	4.724 0819	5.873 1829	7.282 7468	9.007 6079	$1.111\ 3283^{19}$
19	$1.815\ 0420^{19}$	$2.287\ 4502^{19}$	$2.874\ 7685^{19}$	$3.603\ 0432^{19}$	4.503 8040
20	6.534 1512	8.349 1932	$1.063\ 6643^{20}$	$1.351\ 1412^{20}$	$1.711\ 4455^{20}$

n	$m = 91$	$m = 92$	$m = 93$	$m = 94$	$m = 95$
21	2.209 1654 ^{20}	2.862 5805 ^{20}	3.697 4999 ^{20}	4.761 1642 ^{20}	6.112 3054 ^{20}
22	7.029 1627	9.238 3281	1.210 0909 ^{21}	1.579 8408 ^{21}	2.055 9573 ^{21}
23	2.108 7488 ^{21}	2.811 6651 ^{21}	3.735 4979	4.945 5887	6.525 4296
24	5.974 7883	8.083 5371	1.089 5202 ^{22}	1.463 0700 ^{22}	1.957 6289 ^{22}
25	1.601 2433 ^{22}	2.198 7221 ^{22}	3.007 0758	4.096 5960	5.559 6660
26	4.064 6944	5.665 9377	7.864 6598	1.087 1736 ^{23}	1.496 8332 ^{23}
27	9.785 3755	1.385 0070 ^{23}	1.951 6008 ^{23}	2.738 0667	3.825 2403
28	2.236 6572 ^{23}	3.215 1948	4.600 2018	6.551 8025	9.289 8693
29	4.858 9451	7.095 6023	1.031 0797 ^{24}	1.491 0999 ^{24}	2.146 2801 ^{24}
30	1.004 1820 ^{24}	1.490 0765 ^{24}	2.199 6367	3.230 7164	4.721 8163
31	1.975 9710	2.980 1530	4.470 2295	6.669 8662	9.900 5826
32	3.704 9456	5.680 9166	8.661 0696	1.313 1299 ^{25}	1.980 1165 ^{25}
33	6.623 9937	1.032 8939 ^{25}	1.600 9856 ^{25}	2.467 0925	3.780 2224
34	1.129 9754 ^{25}	1.792 3748	2.825 2687	4.426 2543	6.893 3468
35	1.840 2456	2.970 2210	4.762 5958	7.587 8645	1.201 4119 ^{26}
36	2.862 6043	4.702 8500	7.673 0710	1.243 5667 ^{26}	2.002 3531
37	4.255 2226	7.117 8270	1.182 0677 ^{26}	1.949 3748	3.192 9415
38	6.046 8953	1.030 2118 ^{26}	1.741 9945	2.924 0622	4.873 4370
39	8.217 5757	1.426 4471	2.456 6589	4.198 6534	7.122 7156
40	1.068 2848 ^{26}	1.890 0424	3.316 4895	5.773 1484	9.971 8018
41	1.328 8421	2.397 1269	4.287 1694	7.603 6589	1.337 6807 ^{27}
42	1.581 9549	2.910 7970	5.307 9240	9.595 0934	1.719 8752
43	1.802 6928	3.384 6477	6.295 4447	1.160 3369 ^{27}	2.119 8462
44	1.966 5740	3.769 2667	7.153 9144	1.344 9359	2.505 2728
45	2.053 9772	4.020 5512	7.789 8179	1.494 3732	2.839 3092
46	2.053 9772	4.107 9545	8.128 5057	1.591 8324	3.086 2056
47			8.128 5057	1.625 7011	3.217 5335
48					3.217 5335

n	$m = 96$	$m = 97$	$m = 98$	$m = 99$	$m = 100$
1	9.6 ^{1}	9.7 ^{1}	9.8 ^{1}	9.9 ^{1}	1.00 ^{2}
2	4.560 ^{3}	4.656 ^{3}	4.753 ^{3}	4.851 ^{3}	4.950 ^{3}
3	1.428 80 ^{5}	1.474 40 ^{5}	1.520 96 ^{5}	1.568 49 ^{5}	1.617 00 ^{5}
4	3.321 960 ^{6}	3.464 840 ^{6}	3.612 280 ^{6}	3.764 376 ^{6}	3.921 225 ^{6}
5	6.112 4064 ^{7}	6.444 6024 ^{7}	6.791 0864 ^{7}	7.152 3144 ^{7}	7.528 7520 ^{7}
6	9.270 4830 ^{8}	9.881 7237 ^{8}	1.052 6184 ^{9}	1.120 5293 ^{9}	1.192 0524 ^{9}
7	1.191 9192 ^{10}	1.284 6241 ^{10}	1.383 4413 ^{10}	1.488 7032 ^{10}	1.600 7561 ^{10}
8	1.326 0102 ^{11}	1.445 2021 ^{11}	1.573 6645 ^{11}	1.712 0086 ^{11}	1.860 8789 ^{11}
9	1.296 5433 ^{12}	1.429 1443 ^{12}	1.573 6645 ^{12}	1.731 0309 ^{12}	1.902 2318 ^{12}
10	1.127 9926 ^{13}	1.257 6470 ^{13}	1.400 5614 ^{13}	1.557 9279 ^{13}	1.731 0309 ^{13}

n	$m = 96$	$m = 97$	$m = 98$	$m = 99$	$m = 100$
11	8.818 8516 ^{13}	9 946 8442 ^{13}	1.120 4491 ^{14}	1.260 5053 ^{14}	1.416 2980 ^{14}
12	6.246 6865 ^{14}	7.128 5717 ^{14}	8.123 2561	9.243 7052	1.050 4211 ^{15}
13	4.036 3205 ^{15}	4.660 9892 ^{15}	5.373 8464 ^{15}	6.186 1720 ^{15}	7.110 5425
14	2.392 9615 ^{16}	2.796 5935 ^{16}	3.262 6924 ^{16}	3.800 0771 ^{16}	4.418 6943 ^{16}
15	1.308 1523 ^{17}	1.547 4484 ^{17}	1.827 1078 ^{17}	2.153 3770 ^{17}	2.533 3847 ^{17}
16	6.622 5208	7.930 6731	9.478 1215	1.130 5229 ^{18}	1.345 8606 ^{18}
17	3.116 4804 ^{18}	3.778 7325 ^{18}	4.571 7998 ^{18}	5.519 6119	6.650 1349
18	1.367 7886 ^{19}	1.679 4367 ^{19}	2.057 3099 ^{19}	2.514 4899 ^{19}	3.066 4511 ^{19}
19	5.615 1322	6.982 9208	8.662 3575	1.071 9667 ^{20}	1.323 4157 ^{20}
20	2.161 8259 ^{20}	2.723 3391 ^{20}	3.421 6312 ^{20}	4.287 8670	5.359 8337
21	7.823 7509	9.985 5768	1.270 8916 ^{21}	1.613 0547 ^{21}	2.041 8414 ^{21}
22	2.667 1878 ^{21}	3.449 5629 ^{21}	4.448 1206	5.719 0122	7.332 0669
23	8.581 3869	1.124 8575 ^{22}	1.469 8138 ^{22}	1.914 6258 ^{22}	2.486 5270 ^{22}
24	2.610 1718 ^{22}	3.468 3105	4.593 1680	6.062 9817	7.977 6076
25	7.517 2949	1.012 7467 ^{23}	1.359 5777 ^{23}	1.818 8945 ^{23}	2.425 1927 ^{23}
26	2.052 7998 ^{23}	2.804 5292	3.817 2759	5.176 8536	6.995 7482
27	5.322 0734	7.374 8732	1.017 9402 ^{24}	1.399 6678 ^{24}	1.917 3532 ^{24}
28	1.311 5110 ^{24}	1.843 7183 ^{24}	2.581 2056	3.599 1459	4.998 8137
29	3.075 2671	4.386 7780	6.230 4963	8.811 7019	1.241 0848 ^{25}
30	6.868 0965	9.943 3635	1.433 0142 ^{25}	2.056 0638 ^{25}	2.937 2340
31	1.462 2399 ^{25}	2.149 0495 ^{25}	3.143 3859	4.576 4000	6.632 4638
32	2.970 1748	4.432 4147	6.581 4642	9.724 8501	1.430 1250 ^{26}
33	5.760 3390	8.730 5137	1.316 2928 ^{26}	1.974 4393 ^{26}	2.946 9243
34	1.067 3569 ^{26}	1.643 3908 ^{26}	2.516 4422	3.832 7350	5.807 1743
35	1.890 7466	2.958 1035	4.601 4943	7.117 9365	1.095 0672 ^{27}
36	3.203 7650	5.094 5115	8.052 6150	1.265 4109 ^{27}	1.977 2046
37	5.195 2946	8.399 0596	1.349 3571 ^{27}	2.154 6186	3.420 0295
38	8.066 3784	1.326 1673 ^{27}	2.166 0733	3.515 4304	5.670 0490
39	1.199 6153 ^{27}	2.006 2531	3.332 4204	5.498 4937	9.013 9240
40	1.709 4517	2.909 0670	4.915 3201	8.247 7405	1.374 6234 ^{28}
41	2.334 8609	4.044 3126	6.953 3796	1.186 8700 ^{28}	2.011 6440
42	3.057 5560	5.392 4169	9.436 7295	1.639 0109	2.825 8809
43	3.839 7214	6.897 2774	1.228 9694 ^{28}	2.172 6424	3.811 6533
44	4.625 1190	8.464 8404	1.536 2118	2.765 1812	4.937 8236
45	5.344 5820	9.969 7009	1.843 4541	3.379 6659	6.144 8471
46	5.925 5148	1.127 0097 ^{28}	2.123 9798	3.967 4339	7.347 0998
47	6.303 7391	1.222 9254	2.349 9351	4.473 9148	8.441 3487
48	6.435 9670	1.273 8806	2.496 8060	4.846 7411	9.320 6559
49		1.273 8806	2.547 7612	5.044 5672	9.891 3083
50				5.044 5672	1.008 9134 ^{29}

y	$\Phi(y)$	y	$\Phi(y)$	y	$\Phi(y)$
0.00	1.0000	0.40	0.6892	0.80	0.4237
.01	.9920	.41	.6818	.81	.4179
.02	.9840	.42	.6745	.82	.4122
.03	.9761	.43	.6672	.83	.4065
.04	.9681	.44	.6599	.84	.4009
0.05	0.9601	0.45	0.6527	0.85	0.3953
.06	.9522	.46	.6455	.86	.3898
.07	.9442	.47	.6384	.87	.3843
.08	.9362	.48	.6312	.88	.3789
.09	.9283	.49	.6241	.89	.3735
0.10	0.9203	0.50	0.6171	0.90	0.3681
.11	.9124	.51	.6101	.91	.3628
.12	.9045	.52	.6031	.92	.3576
.13	.8966	.53	.5961	.93	.3524
.14	.8887	.54	.5892	.94	.3472
0.15	0.8808	0.55	0.5823	0.95	0.3421
.16	.8729	.56	.5755	.96	.3371
.17	.8650	.57	.5687	.97	.3321
.18	.8572	.58	.5619	.98	.3271
.19	.8493	.59	.5552	.99	.3222
0.20	0.8415	0.60	0.5485	1.00	0.3173
.21	.8337	.61	.5419	.01	.3125
.22	.8259	.62	.5353	.02	.3077
.23	.8181	.63	.5287	.03	.3030
.24	.8103	.64	.5222	.04	.2984
0.25	0.8026	0.65	0.5157	1.05	0.2937
.26	.7949	.66	.5093	.06	.2891
.27	.7872	.67	.5029	.07	.2846
.28	.7795	.68	.4965	.08	.2802
.29	.7718	.69	.4902	.09	.2757
0.30	0.7642	0.70	0.4839	1.10	0.2713
.31	.7566	.71	.4777	.11	.2670
.32	.7490	.72	.4715	.12	.2627
.33	.7414	.73	.4654	.13	.2585
.34	.7339	.74	.4593	.14	.2543
0.35	0.7263	0.75	0.4533	1.15	0.2502
.36	.7189	.76	.4473	.16	.2460
.37	.7114	.77	.4413	.17	.2420
.38	.7039	.78	.4354	.18	.2380
.39	.6965	.79	.4295	.19	.2341

The definition of the Error Function is $\Phi(y) = \sqrt{\frac{2}{\pi}} \int_y^\infty e^{\frac{-v^2}{2}} dy$

y	$\Phi(y)$	y	$\Phi(y)$	y	$\Phi(y)$
1.20	0.2301	1.60	0.1096	2.00	0.0455
.21	.2263	.61	.1074	.01	.0444
.22	.2225	.62	.1052	.02	.0434
.23	.2187	.63	.1031	.03	.0424
.24	.2150	.64	.1010	.04	.0414
1.25	0.2113	1.65	0.0989	2.05	0.0404
.26	.2077	.66	.0969	.06	.0394
.27	.2041	.67	.0949	.07	.0385
.28	.2006	.68	.0930	.08	.0375
.29	.1971	.69	.0910	.09	.0366
1.30	0.1936	1.70	0.0891	2.10	0.0357
.31	.1902	.71	.0873	.11	.0349
.32	.1868	.72	.0854	.12	.0340
.33	.1835	.73	.0836	.13	.0332
.34	.1803	.74	.0819	.14	.0324
1.35	0.1770	1.75	0.0801	2.15	0.0316
.36	.1738	.76	.0784	.16	.0308
.37	.1707	.77	.0767	.17	.0300
.38	.1676	.78	.0751	.18	.0293
.39	.1645	.79	.0735	.19	.0285
1.40	0.1615	1.80	0.0719	2.20	0.0278
.41	.1585	.81	.0703	.21	.0271
.42	.1556	.82	.0688	.22	.0264
.43	.1527	.83	.0673	.23	.0257
.44	.1499	.84	.0658	.24	.0251
1.45	0.1471	1.85	0.0643	2.25	0.0244
.46	.1443	.86	.0629	.26	.0238
.47	.1416	.87	.0615	.27	.0232
.48	.1389	.88	.0601	.28	.0226
.49	.1362	.89	.0588	.29	.0220
1.50	0.1336	1.90	0.0574	2.30	0.0214
.51	.1311	.91	.0561	.31	.0209
.52	.1285	.92	.0549	.32	.0203
.53	.1260	.93	.0536	.33	.0198
.54	.1236	.94	.0524	.34	.0193
1.55	0.1211	1.95	0.0512	2.35	0.0188
.56	.1188	.96	.0500	.36	.0183
.57	.1164	.97	.0488	.37	.0178
.58	.1141	.98	.0477	.38	.0173
.59	.1118	.99	.0466	.39	.0168

y	$\Phi(y)$	y	$\Phi(y)$	y	$\Phi(y)$
2.40	0.0164	2.70	0.0069	3.00	0.0027
.41	.0160	.71	.0067	.10	.0019
.42	.0155	.72	.0065	.20	.0014
.43	.0151	.73	.0063	.30	.0010
.44	.0147	.74	.0061	.40	.0007
2.45	0.0143	2.75	0.0060	.50	0.0005
.46	.0139	.76	.0058	.60	.0003
.47	.0135	.77	.0056	.70	.0002
.48	.0131	.78	.0054	.80	.0001
.49	.0128	.79	.0053	.90	.0001
2.50	0.0124	2.80	0.0051	4.00	0.0001
.51	.0121	.81	.0050		
.52	.0117	.82	.0048		
.53	.0114	.83	.0047		
.54	.0111	.84	.0045		
2.55	0.0108	2.85	0.0044		
.56	.0105	.86	.0042		
.57	.0102	.87	.0041		
.58	.0099	.88	.0040		
.59	.0096	.89	.0039		
2.60	0.0093	2.90	0.0037		
.61	.0091	.91	.0036		
.62	.0088	.92	.0035		
.63	.0085	.93	.0034		
.64	.0083	.94	.0033		
				1.645	0.1
2.65	0.0081	2.95	0.0032	2.576	.01
.66	.0078	.96	.0031	3.291	.001
.67	.0076	.97	.0030	3.890	.0001
.68	.0074	.98	.0029	4.417	.00001
.69	.0071	.99	.0028	4.892	.000001

y	$\phi_{-1}(y)$	$\phi(y)$	$\phi'(y)$	$\phi''(y)$
0.0	+0.50000	+0.39894	−0.00000	−0.39894
0.1	0.53983	0.39695	0.03970	0.39298
0.2	0.57926	0.39104	0.07821	0.37540
0.3	0.61791	0.38139	0.11442	0.34706
0.4	0.65542	0.36827	0.14731	0.30935
0.5	+0.69146	+0.35207	−0.17603	−0.26405
0.6	0.72575	0.33322	0.19993	0.21326
0.7	0.75804	0.31225	0.21858	0.15925
0.8	0.78814	0.28969	0.23175	0.10429
0.9	0.81594	0.26609	0.23948	−0.05056
1.0	+0.84134	+0.24197	−0.24197	+0.00000
1.1	0.86433	0.21785	0.23964	0.04575
1.2	0.88493	0.19419	0.23302	0.08544
1.3	0.90320	0.17137	0.22278	0.11824
1.4	0.91924	0.14973	0.20962	0.14374
1.5	+0.93319	+0.12952	−0.19428	+0.16190
1.6	0.94520	0.11092	0.17747	0.17304
1.7	0.95543	0.09405	0.15988	0.17775
1.8	0.96407	0.07895	0.14211	0.17685
1.9	0.97128	0.06562	0.12467	0.17126
2.0	+0.97725	+0.05399	−0.10798	+0.16197
2.1	0.98214	0.04398	0.09237	0.14998
2.2	0.98610	0.03547	0.07804	0.13622
2.3	0.98928	0.02833	0.06515	0.12152
2.4	0.99180	0.02239	0.05375	0.10660
2.5	+0.99379	+0.01753	−0.04382	+0.09202
2.6	0.99534	0.01358	0.03532	0.07824
2.7	0.99653	0.01042	0.02814	0.06555
2.8	0.99744	0.00792	0.02216	0.05414
2.9	0.99813	0.00595	0.01726	0.04411
3.0	+0.99865	+0.00443	−0.01330	+0.03545
3.1	0.99903	0.00327	0.01013	0.02813
3.2	0.99931	0.00238	0.00763	0.02203
3.3	0.99952	0.00172	0.00568	0.01704
3.4	0.99966	0.00123	0.00419	0.01301
3.5	+0.99977	+0.00087	−0.00305	+0.00982
3.6	0.99984	0.00061	0.00220	0.00732
3.7	0.99989	0.00042	0.00157	0.00539
3.8	0.99993	0.00029	0.00111	0.00392
3.9	0.99995	0.00020	0.00077	0.00282
4.0	+0.99997	+0.00013	−0.00054	+0.00201

The notation is: $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}}$,

y	$\phi'''(y)$	$\phi^{iv}(y)$	$\phi^{v}(y)$	$\phi^{vi}(y)$
0.0	+0.00000	+1.19683	−0.00000	−5.98413
0.1	0.11869	1.16708	0.59146	5.77625
0.2	0.23150	1.07990	1.14197	5.17112
0.3	0.33295	0.94130	1.61420	4.22226
0.4	0.41835	0.76070	1.97770	3.01221
0.5	+0.48409	+0.55010	−2.21141	−1.64481
0.6	0.52783	0.32309	2.30517	−0.23237
0.7	0.54863	+0.09371	2.26012	+1.11354
0.8	0.54694	−0.12468	2.08800	2.29382
0.9	0.52445	0.32034	1.80951	3.23026
1.0	+0.48394	−0.48394	−1.45182	+3.87153
1.1	0.42895	0.60909	1.04580	4.19585
1.2	0.36352	0.69255	0.62301	4.21034
1.3	0.29184	0.73413	−0.21300	3.94753
1.4	0.21800	0.73642	+0.15897	3.45953
1.5	+0.14571	−0.70425	+0.47355	+2.81094
1.6	0.07809	0.64405	0.71813	2.07125
1.7	+0.01759	0.56316	0.88702	1.30785
1.8	−0.03411	0.46915	0.98090	+0.58014
1.9	0.07605	0.36928	1.00583	−0.06467
2.0	−0.10798	−0.26996	+0.97184	−0.59390
2.1	0.13024	0.17646	0.89150	0.98987
2.2	0.14360	0.09274	0.77844	1.24885
2.3	0.14920	−0.02141	0.64604	1.37883
2.4	0.14834	+0.03623	0.50642	1.39654
2.5	−0.14242	+0.07997	+0.36974	−1.32421
2.6	0.13279	0.11053	0.24376	1.18645
2.7	0.12071	0.12926	0.13381	1.00761
2.8	0.10727	0.13793	+0.04287	0.80970
2.9	0.09339	0.13850	−0.02810	0.61102
3.0	−0.07977	+0.13296	−0.07977	−0.42546
3.1	0.06694	0.12313	0.11395	0.26242
3.2	0.05523	0.11066	0.13319	0.12712
3.3	0.04485	0.09690	0.14036	−0.02130
3.4	0.03586	0.08290	0.13840	+0.05607
3.5	−0.02825	+0.06943	−0.13000	+0.10784
3.6	0.02194	0.05703	0.11755	0.13802
3.7	0.01680	0.04599	0.10297	0.15102
3.8	0.01269	0.03646	0.08777	0.15124
3.9	0.00946	0.02842	0.07302	0.14264
4.0	−0.00696	+0.02181	−0.05942	+0.12861

$$\phi_{-1}(y) = \int_{-\infty}^{y} \phi(y)\, dy, \qquad \phi^{t}(y) = \frac{d^{t}\phi(y)}{dy^{t}}.$$

j	$\epsilon = 0.1$	$\epsilon = 0.2$	$\epsilon = 0.3$	$\epsilon = 0.4$	$\epsilon = 0.5$
0	9.0484^{-1}	8.1873^{-1}	7.4082^{-1}	6.7032^{-1}	6.0653^{-1}
1	9.0484^{-2}	1.6375	2.2225	2.6813	3.0327
2	4.5242^{-3}	1.6375^{-2}	3.3337^{-2}	5.3626^{-2}	7.5816^{-2}
3	1.5081^{-4}	1.0916^{-3}	3.3337^{-3}	7.1501^{-3}	1.2636
4	3.7702^{-6}	5.4582^{-5}	2.5003^{-4}	7.1501^{-4}	1.5795^{-3}
5		2.1833^{-6}	1.5002^{-5}	5.7201^{-5}	1.5795^{-4}
6				3.8134^{-6}	1.3163^{-5}

j	$\epsilon = 0.6$	$\epsilon = 0.7$	$\epsilon = 0.8$	$\epsilon = 0.9$
0	5.4881^{-1}	4.9659^{-1}	4.4933^{-1}	4.0657^{-1}
1	3.2929	3.4761	3.5946	3.6591
2	9.8786^{-2}	1.2166	1.4379	1.6466
3	1.9757	2.8388^{-2}	3.8343^{-2}	4.9398^{-2}
4	2.9636^{-3}	4.9679^{-3}	7.6685^{-3}	1.1115
5	3.5563^{-4}	6.9551^{-4}	1.2270	2.0006^{-3}
6	3.5563^{-5}	8.1143^{-5}	1.6360^{-4}	3.0009^{-4}
7	3.0483^{-6}	8.1143^{-6}	1.8697^{-5}	3.8584^{-5}
8			1.8697^{-6}	4.3406^{-6}

j	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 3$	$\epsilon = 4$	$\epsilon = 5$
0	3.6788^{-1}	1.3534^{-1}	4.9787^{-2}	1.8316^{-2}	6.7379^{-3}
1	3.6788	2.7067	1.4936^{-1}	7.3263	3.3690^{-2}
2	1.8394	2.7067	2.2404	1.4653^{-1}	8.4224
3	6.1313^{-2}	1.8045	2.2404	1.9537	1.4037^{-1}
4	1.5328	9.0224^{-2}	1.6803	1.9537	1.7547

j	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 3$	$\epsilon = 4$	$\epsilon = 5$
5	3.0657^{-3}	3.6089^{-2}	1.0082^{-1}	1.5629^{-1}	1.7547^{-1}
6	5.1094^{-4}	1.2030	5.0409^{-2}	1.0420	1.4622
7	7.2992^{-5}	3.4371^{-3}	2.1604	5.9540^{-2}	1.0444
8	9.1240^{-6}	8.5927^{-4}	8.1015^{-3}	2.9770	6.5278^{-2}
9	1.0138	1.9095	2.7005	1.3231	3.6266
10		3.8190^{-5}	8.1015^{-4}	5.2925^{-3}	1.8133
11		6.9436^{-6}	2.2095	1.9245	8.2422^{-3}
12		1.1573	5.5238^{-5}	6.4151^{-4}	3.4342
13			1.2747	1.9739	1.3209
14			2.7315^{-6}	5.6397^{-5}	4.7174^{-4}
15				1.5039	1.5725
16				3.7598^{-6}	4.9139^{-5}
17					1.4453
18					4.0146^{-6}
19					1.0565

j	$\epsilon = 6$	$\epsilon = 7$	$\epsilon = 8$	$\epsilon = 9$	$\epsilon = 10$
0	2.4788^{-3}	9.1188^{-4}	3.3546^{-4}	1.2341^{-4}	4.5400^{-5}
1	1.4873^{-2}	6.3832^{-3}	2.6837^{-3}	1.1107^{-3}	4.5400^{-4}
2	4.4618	2.2341^{-2}	1.0735^{-2}	4.9981	2.2700^{-3}
3	8.9235	5.2129	2.8626	1.4994^{-2}	7.5667
4	1.3385^{-1}	9.1226	5.7252	3.3737	1.8917^{-2}
5	1.6062	1.2772^{-1}	9.1604	6.0727	3.7833
6	1.6062	1.4900	1.2214^{-1}	9.1090	6.3055
7	1.3768	1.4900	1.3959	1.1712^{-1}	9.0079
8	1.0326	1.3038	1.3959	1.3176	1.1260^{-1}
9	6.8838^{-2}	1.0140	1.2408	1.3176	1.2511
10	4.1303	7.0983^{-2}	9.9262^{-2}	1.1858	1.2511
11	2.2529	4.5171	7.2190	9.7020^{-2}	1.1374
12	1.1264	2.6350	4.8127	7.2765	9.4780^{-2}
13	5.1990^{-3}	1.4188	2.9616	5.0376	7.2908
14	2.2281	7.0942^{-3}	1.6924	3.2384	5.2077
15	8.9126^{-4}	3.3106	9.0260^{-3}	1.9431	3.4718
16	3.3422	1.4484	4.5130	1.0930	2.1699
17	1.1796	5.9640^{-4}	2.1238	5.7863^{-3}	1.2764
18	3.9320^{-5}	2.3193	9.4389^{-4}	2.8932	7.0911^{-3}
19	1.2417	8.5449^{-5}	3.9743	1.3704	3.7322

j	$\epsilon = 6$	$\epsilon = 7$	$\epsilon = 8$	$\epsilon = 9$	$\epsilon = 10$
20	3.7251^{-6}	2.9907^{-5}	1.5897^{-4}	6.1670^{-4}	1.8661^{-3}
21	1.0643	9.9690^{-6}	6.0561^{-5}	2.6430	8.8861^{-4}
22		3.1720	2.2022	1.0812	4.0391
23			7.6598^{-6}	4.2309^{-5}	1.7561
24			2.5533	1.5866	7.3173^{-5}
25				5.7117^{-6}	2.9269
26				1.9771	1.1257
27					4.1694^{-6}
28					1.4891

j	$\epsilon = 11$	$\epsilon = 12$	$\epsilon = 13$	$\epsilon = 14$	$\epsilon = 15$
0	1.6702^{-5}	6.1442^{-6}	2.2603^{-6}		
1	1.8372^{-4}	7.3731^{-5}	2.9384^{-5}	1.1641^{-5}	4.5885^{-6}
2	1.0105^{-3}	4.4238^{-4}	1.9100^{-4}	8.1490	3.4414^{-5}
3	3.7050	1.7695^{-3}	8.2766	3.8029^{-4}	1.7207^{-4}
4	1.0189^{-2}	5.3086	2.6899^{-3}	1.3310^{-3}	6.4526
5	2.2415	1.2741^{-2}	6.9937	3.7268	1.9358^{-3}
6	4.1095	2.5481	1.5153^{-2}	8.6959	4.8395
7	6.4577	4.3682	2.8141	1.7392	1.0370^{-2}
8	8.8794	6.5523	4.5730	3.0436	1.9444
9	1.0853^{-1}	8.7364	6.6054	4.7344	3.2407
10	1.1938	1.0484^{-1}	8.5870	6.6282	4.8611
11	1.1938	1.1437	1.0148^{-1}	8.4359	6.6287
12	1.0943	1.1437	1.0994	9.8418	8.2859
13	9.2595^{-2}	1.0537	1.0994	1.0599^{-1}	9.5607
14	7.2753	9.0489^{-2}	1.0209	1.0599	1.0244^{-1}
15	5.3352	7.2391	8.8475^{-2}	9.8923^{-2}	1.0244
16	3.6680	5.4293	7.1886	8.6558	9.6034^{-2}
17	2.3734	3.8325	5.4972	7.1283	8.4736
18	1.4504	2.5550	3.9702	5.5442	7.0613
19	8.3971^{-3}	1.6137	2.7164	4.0852	5.5747
20	4.6184	9.6820^{-3}	1.7657	2.8597	4.1810
21	2.4192	5.5326	1.0930	1.9064	2.9865
22	1.2096	3.0178	6.4589^{-3}	1.2132	2.0362
23	5.7849^{-4}	1.5745	3.6507	7.3846^{-3}	1.3280
24	2.6514	7.8725^{-4}	1.9775	4.3077	8.2998^{-3}

j	$\epsilon = 11$	$\epsilon = 12$	$\epsilon = 13$	$\epsilon = 14$	$\epsilon = 15$
25	1.1666^{-4}	3.7788^{-4}	1.0283^{-3}	2.4123^{-3}	4.9799^{-3}
26	4.9357^{-5}	1.7441	5.1414^{-4}	1.2989	2.8730
27	2.0109	7.7513^{-5}	2.4755	6.7352^{-4}	1.5961
28	7.8998^{-6}	3.3220	1.1493	3.3676	8.5506^{-4}
29	2.9965	1.3746	5.1522^{-5}	1.6257	4.4227
30	1.0987	5.4985^{-6}	2.2326	7.5868^{-5}	2.2114
31		2.1284	9.3625^{-6}	3.4263	1.0700
32			3.8035	1.4990	5.0157^{-5}
33			1.4984	6.3594^{-6}	2.2799
34				2.6186	1.0058
35				1.0474	4.3107^{-6}
36					1.7961

j	$\epsilon = 16$	$\epsilon = 17$	$\epsilon = 18$	$\epsilon = 19$	$\epsilon = 20$
0					
1	1.8006^{-6}				
2	1.4405^{-5}	5.9822^{-6}	2.4673^{-6}	1.0113^{-6}	
3	7.6824	3.3899^{-5}	1.4804^{-5}	6.4049	2.7482^{-6}
4	3.0730^{-4}	1.4407^{-4}	6.6616	3.0423^{-5}	1.3741^{-5}
5	9.8335	4.8984	2.3982^{-4}	1.1561^{-4}	5.4964
6	2.6223^{-3}	1.3879^{-3}	7.1945	3.6610	1.8321^{-4}
7	5.9937	3.3706	1.8500^{-3}	9.9369	5.2347
8	1.1987^{-2}	7.1625	4.1625	2.3600^{-3}	1.3087^{-3}
9	2.1311	1.3529^{-2}	8.3251	4.9822	2.9082
10	3.4098	2.3000	1.4985^{-2}	9.4662	5.8163
11	4.9597	3.5545	2.4521	1.6351^{-2}	1.0575^{-2}
12	6.6129	5.0355	3.6782	2.5889	1.7625
13	8.1389	6.5849	5.0929	3.7837	2.7116
14	9.3016	7.9960	6.5480	5.1351	3.8737
15	9.9218	9.0621	7.8576	6.5044	5.1649
16	9.9218	9.6285	8.8397	7.7240	6.4561
17	9.3381	9.6285	9.3597	8.6327	7.5954
18	8.3006	9.0935	9.3597	9.1123	8.4394
19	6.9899	8.1363	8.8671	9.1123	8.8835

j	$\epsilon = 16$	$\epsilon = 17$	$\epsilon = 18$	$\epsilon = 19$	$\epsilon = 20$
20	5.5920^{-2}	6.9159^{-2}	7.9804^{-2}	8.6567^{-2}	8.8835^{-2}
21	4.2605	5.5986	6.8403	7.8323	8.4605
22	3.0986	4.3262	5.5966	6.7642	7.6914
23	2.1555	3.1976	4.3800	5.5878	6.6881
24	1.4370	2.2650	3.2850	4.4237	5.5735
25	9.1969^{-3}	1.5402	2.3652	3.3620	4.4588
26	5.6596	1.0070	1.6374	2.4569	3.4298
27	3.3539	6.3406^{-3}	1.0916	1.7289	2.5406
28	1.9165	3.8497	7.0176^{-3}	1.1732	1.8147
29	1.0574	2.2567	4.3558	7.6864^{-3}	1.2515
30	5.6393^{-4}	1.2788	2.6135	4.8680	8.3435^{-3}
31	2.9106	7.0128^{-4}	1.5175	2.9836	5.3829
32	1.4553	3.7255	8.5359^{-4}	1.7715	3.3643
33	7.0561^{-5}	1.9192	4.6559	1.0200	2.0390
34	3.3205	9.5961^{-5}	2.4649	5.6998^{-4}	1.1994
35	1.5179	4.6609	1.2677	3.0942	6.8537^{-4}
36	6.7464^{-6}	2.2010	6.3383^{-5}	1.6330	3.8076
37	2.9174	1.0113	3.0835	8.3859^{-5}	2.0582
38	1.2284	4.5241^{-6}	1.4606	4.1930	1.0833
39		1.9720	6.7413^{-6}	2.0427	5.5551^{-5}
40			3.0336	9.7030^{-6}	2.7776
41			1.3318	4.4965	1.3549
42				2.0341	6.4520^{-6}
43					3.0009
44					1.3641

ν	$\epsilon = 0.1$	$\epsilon = 0.2$	$\epsilon = 0.3$	$\epsilon = 0.4$	$\epsilon = 0.5$
0	1.0000	1.0000	1.0000	1.0000	1.0000
1	9.5163^{-2}	1.8127^{-1}	2.5918^{-1}	3.2968^{-1}	3.9347^{-1}
2	4.6788^{-3}	1.7523^{-2}	3.6936^{-2}	6.1552^{-2}	9.0204^{-2}
3	1.5465^{-4}	1.1485^{-3}	3.5995^{-3}	7.9263^{-3}	1.4388
4	3.8468^{-6}	5.6840^{-5}	2.6581^{-4}	7.7625^{-4}	1.7516^{-3}
5		2.2582^{-6}	1.5785^{-5}	6.1243^{-5}	1.7212^{-4}
6				4.0427^{-6}	1.4165^{-5}
7					1.0024^{-6}

ν	$\epsilon = 0.6$	$\epsilon = 0.7$	$\epsilon = 0.8$	$\epsilon = 0.9$
0	1.0000	1.0000	1.0000	1.0000
1	4.5119^{-1}	5.0341^{-1}	5.5067^{-1}	5.9343^{-1}
2	1.2190	1.5580	1.9121	2.2752
3	2.3115^{-2}	3.4142^{-2}	4.7423^{-2}	6.2857^{-2}
4	3.3581^{-3}	5.7535^{-3}	9.0799^{-3}	1.3459
5	3.9449^{-4}	7.8554^{-4}	1.4113	2.3441^{-3}
6	3.8856^{-5}	9.0026^{-5}	1.8434^{-4}	3.4349^{-4}
7	3.2931^{-6}	8.8836^{-6}	2.0747^{-5}	4.3401^{-5}
8			2.0502^{-6}	4.8172^{-6}

ν	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 3$	$\epsilon = 4$	$\epsilon = 5$
0	1.0000	1.0000	1.0000	1.0000	1.0000
1	6.3212^{-1}	8.6466^{-1}	9.5021^{-1}	9.8168^{-1}	9.9326^{-1}
2	2.6424	5.9399	8.0085	9.0842	9.5957
3	8.0301^{-2}	3.2332	5.7681	7.6190	8.7535
4	1.8988	1.4288	3.5277	5.6653	7.3497

ν	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 3$	$\epsilon = 4$	$\epsilon = 5$
5	3.6598^{-3}	5.2653^{-2}	1.8474^{-1}	3.7116^{-1}	5.5951^{-1}
6	5.9418^{-4}	1.6564	8.3918^{-2}	2.1487	3.8404
7	8.3241^{-5}	4.5338^{-3}	3.3509	1.1067	2.3782
8	1.0249	1.0967	1.1905	5.1134^{-2}	1.3337
9	1.1252^{-6}	2.3745^{-4}	3.8030^{-3}	2.1363	6.8094^{-2}
10		4.6498^{-5}	1.1025	8.1322^{-3}	3.1828
11		8.3082^{-6}	2.9234^{-4}	2.8398	1.3695
12		1.3646	7.1387^{-5}	9.1523^{-4}	5.4531^{-3}
13			1.6149	2.7372	2.0189
14			3.4019^{-6}	7.6328^{-5}	6.9799^{-4}
15				1.9932	2.2625
16				4.8926^{-6}	6.9008^{-5}
17				1.1328	1.9869
18					5.4163^{-6}
19					1.4017

ν	$\epsilon = 6$	$\epsilon = 7$	$\epsilon = 8$	$\epsilon = 9$	$\epsilon = 10$
0	1.0000	1.0000	1.0000	1.0000	1.0000
1	9.9752^{-1}	9.9909^{-1}	9.9966^{-1}	9.9988^{-1}	9.9995^{-1}
2	9.8265	9.9270	9.9698	9.9877	9.9950
3	9.3803	9.7036	9.8625	9.9377	9.9723
4	8.4880	9.1823	9.5762	9.7877	9.8966
5	7.1494	8.2701	9.0037	9.4504	9.7075
6	5.5432	6.9929	8.0876	8.8431	9.3291
7	3.9370	5.5029	6.8663	7.9322	8.6986
8	2.5602	4.0129	5.4704	6.7610	7.7978
9	1.5276	2.7091	4.0745	5.4435	6.6718
10	8.3924^{-2}	1.6950	2.8338	4.1259	5.4207
11	4.2621	9.8521^{-2}	1.8411	2.9401	4.1696
12	2.0092	5.3350	1.1192	1.9699	3.0322
13	8.8275^{-3}	2.7000	6.3797^{-2}	1.2423	2.0844
14	3.6285	1.2811	3.4181	7.3851^{-2}	1.3554
15	1.4004	5.7172^{-3}	1.7257	4.1466	8.3458^{-2}
16	5.0910^{-4}	2.4066	8.2310^{-3}	2.2036	4.8740
17	1.7488	9.5818^{-4}	3.7180	1.1106	2.7042
18	5.6917^{-5}	3.6178	1.5943	5.3196^{-3}	1.4278
19	1.7597	1.2985	6.5037^{-4}	2.4264	7.1865^{-3}

ν	$\epsilon = 6$	$\epsilon = 7$	$\epsilon = 8$	$\epsilon = 9$	$\epsilon = 10$
20	5.1802^{-6}	4.4422^{-5}	2.5294^{-3}	1.0560^{-3}	3.4543^{-4}
21	1.4551	1.4495	9.3968^{-5}	4.3925^{-4}	1.5883
22		4.5263^{-6}	3.3407	1.7495	6.9965^{-4}
23		1.3543	1.1385	6.6828^{-5}	2.9574
24			3.7255^{-6}	2.4519	1.2012
25			1.1722	8.6531^{-6}	4.6949^{-5}
26				2.9414	1.7680
27					6.4229^{-6}
28					2.2535

ν	$\epsilon = 11$	$\epsilon = 12$	$\epsilon = 13$	$\epsilon = 14$	$\epsilon = 15$
0	1.0000	1.0000			
1	9.9998^{-1}	9.9999^{-1}	1.0000	1.0000	
2	9.9980	9.9992	9.9997^{-1}	9.9999^{-1}	1.0000
3	9.9879	9.9948	9.9978	9.9991	9.9996^{-1}
4	9.9508	9.9771	9.9895	9.9953	9.9979
5	9.8490	9.9240	9.9626	9.9819	9.9914
6	9.6246	9.7966	9.8927	9.9447	9.9721
7	9.2139	9.5418	9.7411	9.8577	9.9237
8	8.5681	9.1050	9.4597	9.6838	9.8200
9	7.6801	8.4497	9.0024	9.3794	9.6255
10	6.5949	7.5761	8.3419	8.9060	9.3015
11	5.4011	6.5277	7.4832	8.2432	8.8154
12	4.2073	5.3840	6.4684	7.3996	8.1525
13	3.1130	4.2403	5.3690	6.4154	7.3239
14	2.1871	3.1846	4.2696	5.3555	6.3678
15	1.4596	2.2798	3.2487	4.2956	5.3435
16	9.2624^{-2}	1.5558	2.3639	3.3064	4.3191
17	5.5924	1.0129	1.6451	2.4408	3.3588
18	3.2191	6.2966^{-2}	1.0954	1.7280	2.5114
19	1.7687	3.7416	6.9833^{-2}	1.1736	1.8053
20	9.2895^{-3}	2.1280	4.2669	7.6505^{-2}	1.2478
21	4.6711	1.1598	2.5012	4.7908	8.2972^{-2}
22	2.2519	6.0651^{-3}	1.4081	2.8844	5.3106
23	1.0423	3.0474	7.6225^{-3}	1.6712	3.2744
24	4.6386^{-4}	1.4729	3.9718	9.3276^{-3}	1.9465

ν	$\epsilon = 11$	$\epsilon = 12$	$\epsilon = 13$	$\epsilon = 14$	$\epsilon = 15$
25	1.9871^{-4}	6.8563^{-4}	1.9943^{-3}	5.0199^{-3}	1.1165^{-2}
26	8.2050^{-5}	3.0776	9.6603^{-4}	2.6076	6.1849^{-3}
27	3.2693	1.3335	4.5190	1.3087	3.3119
28	1.2584	5.5836^{-5}	2.0435	6.3513^{-4}	1.7158
29	4.6847^{-6}	2.2616	8.9416^{-5}	2.9837	8.6072^{-4}
30	1.6882	8.8701^{-6}	3.7894	1.3580	4.1845
31		3.3716	1.5568	5.9928^{-5}	1.9731
32		1.2432	6.2052^{-6}	2.5665	9.0312^{-5}
33			2.4017	1.0675	4.0155
34				4.3154^{-6}	1.7356
35				1.6968	7.2978^{-6}
36					2.9871
37					1.1910

ν	$\epsilon = 16$	$\epsilon = 17$	$\epsilon = 18$	$\epsilon = 19$	$\epsilon = 20$
0					
1					
2	1.0000	1.0000			
3	9.9998^{-1}	9.9999^{-1}	1.0000	1.0000	
4	9.9991	9.9996	9.9998^{-1}	9.9999^{-1}	1.0000
5	9.9960	9.9982	9.9992	9.9996	9.9998^{-1}
6	9.9862	9.9933	9.9968	9.9985	9.9993
7	9.9599	9.9794	9.9896	9.9948	9.9974
8	9.9000	9.9457	9.9711	9.9849	9.9922
9	9.7801	9.8740	9.9294	9.9613	9.9791
10	9.5670	9.7388	9.8462	9.9114	9.9500
11	9.2260	9.5088	9.6963	9.8168	9.8919
12	8.7301	9.1533	9.4511	9.6533	9.7861
13	8.0688	8.6498	9.0833	9.3944	9.6099
14	7.2545	7.9913	8.5740	9.0160	9.3387
15	6.3247	7.1917	7.9192	8.5025	8.9514
16	5.3326	6.2855	7.1335	7.8521	8.4349
17	4.3404	5.3226	6.2495	7.0797	7.7893
18	3.4066	4.3598	5.3135	6.2164	7.0297
19	2.5765	3.4504	4.3776	5.3052	6.1858

ν	$\epsilon = 16$	$\epsilon = 17$	$\epsilon = 18$	$\epsilon = 19$	$\epsilon = 20$
20	1.8775^{-1}	2.6368^{-1}	3.4908^{-1}	4.3939^{-1}	5.2974^{-1}
21	1.3183	1.9452	2.6928	3.5283	4.4091
22	8.9227^{-2}	1.3853	2.0088	2.7450	3.5630
23	5.8241	9.5272^{-2}	1.4491	2.0687	2.7939
24	3.6686	6.3296	1.0111	1.5098	2.1251
25	2.2315	4.0646	6.8260^{-2}	1.0675	1.5677
26	1.3119	2.5245	4.4608	7.3126^{-2}	1.1218
27	7.4589^{-3}	1.5174	2.8234	4.8557	7.7887^{-2}
28	4.1051	8.8335^{-3}	1.7318	3.1268	5.2481
29	2.1886	4.9838	1.0300	1.9536	3.4334
30	1.1312	2.7272	5.9443^{-3}	1.1850	2.1818
31	5.6726^{-4}	1.4484	3.3308	6.9819^{-3}	1.3475
32	2.7620	7.4708^{-4}	1.8133	3.9982	8.0918^{-3}
33	1.3067	3.7453	9.5975^{-4}	2.2267	4.7274
34	6.0108^{-5}	1.8260	4.9416	1.2067	2.6884
35	2.6903	8.6644^{-5}	2.4767	6.3674^{-4}	1.4890
36	1.1724	4.0035	1.2090	3.2732	8.0366^{-4}
37	4.9772^{-6}	1.8025	5.7519^{-5}	1.6401	4.2290
38	2.0599	7.9123^{-6}	2.6684	8.0154^{-5}	2.1708
39		3.3882	1.2078	3.8224	1.0875
40		1.4162	5.3365^{-6}	1.7797	5.3202^{-5}
41			2.3030	8.0940^{-6}	2.5426
42				3.5975	1.1877
43				1.5634	5.4252^{-6}
44					2.4243
45					1.0603

s'	$P = .99$	$P = .98$	$P = .95$	$P = .90$	$P = .80$	$P = .70$
1	0.000157	0.000628	0.00393	0.0158	0.0642	0.148
2	0.0201	0.0404	0.103	0.211	0.446	0.713
3	0.115	0.185	0.352	0.584	1.005	1.424
4	0.297	0.429	0.711	1.064	1.649	2.195
5	0.554	0.752	1.145	1.610	2.343	3.000
6	0.872	1.134	1.635	2.204	3.070	3.828
7	1.239	1.564	2.167	2.833	3.822	4.671
8	1.646	2.032	2.733	3.490	4.594	5.527
9	2.088	2.532	3.325	4.168	5.380	6.393
10	2.558	3.059	3.940	4.865	6.179	7.267
11	3.053	3.609	4.575	5.578	6.989	8.148
12	3.571	4.178	5.226	6.304	7.807	9.034
13	4.107	4.765	5.892	7.042	8.634	9.926
14	4.660	5.368	6.571	7.790	9.467	10.821
15	5.229	5.985	7.261	8.547	10.307	11.721
16	5.812	6.614	7.962	9.312	11.152	12.624
17	6.408	7.255	8.672	10.085	12.002	13.531
18	7.015	7.906	9.390	10.865	12.857	14.440
19	7.633	8.567	10.117	11.651	13.716	15.352
20	8.260	9.237	10.851	12.443	14.578	16.266
21	8.897	9.915	11.591	13.240	15.445	17.182
22	9.542	10.600	12.338	14.041	16.314	18.101
23	10.196	11.293	13.091	14.848	17.187	19.021
24	10.856	11.992	13.848	15.659	18.062	19.943
25	11.524	12.697	14.611	16.473	18.940	20.867
26	12.198	13.409	15.379	17.292	19.820	21.792
27	12.879	14.125	16.151	18.114	20.703	22.719
28	13.565	14.847	16.928	18.939	21.588	23.647
29	14.256	15.574	17.708	19.768	22.475	24.577
30	14.953	16.306	18.493	20.599	23.364	25.508

Taken from *Statistical Methods for Research Workers*, by R. A. Fisher. Published by Oliver & Boyd, Edinburgh.

s'	$P = .50$	$P = .30$	$P = .20$	$P = .10$	$P = .05$	$P = .02$	$P = .01$
1	0.455	1.074	1.642	2.706	3.841	5.412	6.635
2	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	2.366	3.665	4.642	6.251	7.815	9.837	11.341
4	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	11.340	14.011	15.812	18.549	21.026	24.054	26.217
13	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	17.338	20.601	22.760	25.989	28.869	32.346	34.805
19	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	23.337	27.096	29.553	33.196	36.415	40.270	42.980
25	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	29.336	33.530	36.250	40.256	43.773	47.962	50.892

For larger values of s' use Appendix V, with $y = \sqrt{2s' - 1} - \sqrt{2\chi^2}$ and $P = \phi_{-1}(y)$.

IX. STANDARD DEVIATIONS OF IMPORTANT STATISTICS *

(N is the number of observations from which the statistic is computed.)

Statistic		Standard Deviation of Statistic				
				Special formula for		
Name	Symbol	Symbol	General Formula	Normal Law	Poisson Law	Binomial Law
Average	$\bar{n}$	$\sigma(\bar{n})$	$\frac{\sigma}{\sqrt{N}}$	$\frac{\sigma}{\sqrt{N}}$	$\sqrt{\frac{\epsilon}{N}}$	$\sqrt{\frac{mp(1-p)}{N}}$
Standard Deviation	σ	$\sigma(\sigma)$	$\sqrt{\frac{\mu_4(d)-[\mu_2(d)]^2}{4N\mu_2(d)}}$	$\frac{\sigma}{\sqrt{2N}}$	$\sqrt{\frac{2\epsilon+1}{4N}}$	$\sqrt{\frac{2(m-1)p(1-p)+(2p-1)^2}{4N}}$
Asymmetry (Skewness)	$\sqrt{\beta_1}$	$\sigma(\sqrt{\beta_1})$		$\sqrt{\frac{6}{N}}$		
Flatness (Kurtosis)	β_2	$\sigma(\beta_2)$		$\sqrt{\frac{24}{N}}$		

* Taken from *Statistical Methods for Research Workers* by R. A. Fisher. Published by Oliver & Boyd, Edinburgh.

X. CRITERIA FOR CHOICE OF DISTRIBUTION CURVES *

	$J<0$	$J=0$	$J>0$
$\beta_1=0$	IV	Normal	II
$\beta_1>0$	IV	III	I Binomial Poisson

* The Gram-Charlier Series can represent any of these types.

INDEX

Addition Theorem, 5, 12
Asymmetry
  definition of, 298
Average
  definition of, 177, 183
Axioms, Fundamental, 4

Bad Penny, The, 125–127
Bayes' Theorem, 117–132, 265–266
  statement of, 121
  substitute for, 268–270
  uses of, 127–131
Bernoulli's Theorem, 82–116
  proof of, 102–103, 108–111
  statement of, 95, 100
Binomial Coefficients
  table of, 439–452
  (see also *Combinations*)
Binomial Law
  application to traffic problems, 334–336, 345, 347
  as empirical distribution function, 304–305
  Gram-Charlier approximations for, 206–213, 255–261
  relation to Normal Law, 208–211
  relation to Pearson's Curves, 245–246
  relation to Poisson Law, 214–216
  relation to problem of independent trials, 63, 91–101
  standard deviation of statistics of, 314, 470
  statistics of, 471
Binomial Theorem, 29–31

Certainty, 3, 4, 87–88
Change of Variable, 150–163, 261–263
Channel
  definition of, 323
Cogent Reason
  doctrine of, 6, 117–119
Collision of Molecules
  change of velocity on, 171–173, 390–393, 395–398
Combinations, 12–38
  definition of, 15
  fundamental formulae, 26, 28, 31, 60, 69
  tables of, 439–452
Complete Group, 8
Composition of Events
  laws of, 12–14
Congestion
  problems of, 321–388
Continuous Variables, 133–176, 183–186
Control Charts, 315–317
Conventions, Fundamental, 4
Curve Fitting, 265–315

Delays, 372–388
  at cooperative channels, 378–387
  at non-cooperative channels, 376–378
  effect upon traffic congestion, 374–376, 380–382
Deviation, 188–191
Distribution Curve, 96–97
  (see also *Distribution Functions*)
Distribution Functions, 96–97
  criteria for choice of, 296–302, 470, 471
  derived empirically, 144–146, 297–310
  empirical, 241–255
  for continuous variables, 141–146
  many variables, 147–150
  most frequently used, 205–264
  transformation of, 150–163, 261–263

Distribution Functions, (*continued*)
type criterion for, 300
(see also *Binomial Law*, *Engset Law*, *Erlang Law*, *Exponential Law*, *Gram-Charlier Series*, *Normal Law*, *Pearson's Curves*, and *Poisson Law*)
Divergence
definition, of, 290

Engset Law
application to traffic problems, 340–342, 345, 347, 351–354
computation charts for, 352–353
Equally Likely, 5, 56
Erlang Law
application to traffic problems, 342–343, 345, 347
Error Function
(see *Normal Law*)
Exponential Law, 378–387
Expectation
definition of, 177–186
mathematical, 179
moral, 179, 195–196
of a probability, 199–202, 401–403

Factorials, 20–25, 103–107
tables of, 427–438
Flatness
definition of, 299
Fluctuation Phenomena, 389–423

Gamma Function
(see *Factorials*)
Gas Molecules
change of velocity on collision, 171–173, 390–393, 395–398
distribution of speed, 167
distribution of velocity, 163–171
in spherical coordinates, 166
Goodness of Fit, 268–297
relation to Bayes' Theorem, 268–270
the χ^2 criterion, 280–291
table of, 468–469
Gram-Charlier Series, 251–261
as an empirical distribution function, 307–310
for Binomial Law, 206–213, 255–261
for Poisson Law, 237–240, 261
table of Normal Law and its derivatives, 456–457

H-Function, The, 400–403
Hermite Polynomials, 252–255
Hunting Problems, 356–369

Impossibility, 4, 138–140
Independent Trials, 62–65, 82–116
Insufficient Reason
doctrine of, 6, 117–119

Jacobian, 153–163
general significance of, 173–174

Kinetic Theory of Gases, 163–171, 390–417
density fluctuations, 410–414
diffusion, 414–417
fundamental equation of, 399–400
H-Function, the, 400–403
Maxwell's Law of velocity, 163–165, 404–406
mean free path, 409–410
number of collisions, 409
pressure, 406–408
Kurtosis
(see *Flatness*)

Limit, 83–86
definition of, 85
probability as a, 88–91
Line Segment
choice of point on, 133–141

Maxwell's Equation, 163–171
change of variable in, 165–171
derivation of, 163–165, 390–406
Median
definition of, 187–188
Moment
definition of, 183

Multiplication Theorem, 12, 48–52, 113–116
Mutually Exclusive Events, 7

Normal Law
in one variable, 169
in several variables, 165, 280–285
logical standing of, 205, 241–244
relation to Binomial Law, 208–211
relation to Pearson's Curves, 246
relation to Poisson Law, 238
standard deviation of statistics of, 314, 470
statistics of, 471
table of, and its derivatives, 456–457
table of Error Function, 453–455

Paradoxes
Bertrand's "Box," 121–122, 131–132
"Life on Mars," 117–119
"of the impossible," 138–140
"St. Petersburg," 194–199
Pascal's Triangle, 27–29
Pearson's Curves, 244–251
Permutations, 12–38
definition of, 15
fundamental formulae, 25, 34, 36
Poisson Law
application to traffic problems, 227–229, 233–237, 336–338, 345, 347, 354–355
application to variable traffic density, 233–237
application to warehouse problem, 227–232
as empirical distribution function, 305–307
Gram-Charlier approximation to, 237–240, 261
relation to Binomial Law, 214–216
relation to Normal Law, 238
relation to random distributions, 220–227
standard deviation of statistics of, 314, 470
statistics of, 471
tables of, 458–467
Population
definition of, 267
Probability
alternative compound, 53–81, 149
as a statistical ratio, 9, 82–113
complementary, 39
compound, 48–52, 113–116, 149
conditional, 43–48, 148–149
definition of, 1–11
determined by experiment, 112, 119, 125–132
elementary principles of, 39–81
fundamental theorems, 8, 48, 54, 63, 65, 68
irrational values of, 135–137
measure of, 7
unconditional, 39
unit of measure for, 3, 4
Psychic Research
examples from, 41–42, 44, 45, 46, 51, 57–60

Random
"collectively at random," definition of, 218
"individually at random," definition of, 218
non-random distribution, examples of, 143–146
Poisson Law appropriate to random distribution, 220–227
"random distribution," definition of, 141
Repeated Trials, 62–70, 91–101
Root Mean Square
definition of, 183

Schottky Effect, 417–423
Schroteffekt
(see *Schottky Effect*)
Sets of Numbers, 84–88
bounded set, 86–88
closed set, 86
open set, 86
Sheppard's Corrections, 310–312
Skewness
(see *Asymmetry*)

Source
definition of, 323
Standard Deviation
definition of, 190
of important statistics, 312–315, 470
Statistics
definition of, 312
of common distributions, 471
distribution of, 312–314
Stirling's Formula, 103–107
Switch
definition of, 323

Traffic Problems, 294–295, 321–388
application of Binomial Law to, 334–336, 345, 347
application of Engset Law to, 340–342, 345, 347, 351–354
application of Erlang Law to, 342–343, 345, 347
Traffic Problems, (*continued*)
application of Poisson Law to, 227–229, 336–338, 345, 347, 354–355
delays, 372–388
double connections, 369–372
formulae for loss in
computation charts for, 351–355
lost calls cleared, 339–344
lost calls held, 329–339
numerical comparison of, 346–351
recapitulation of, 345, 347
hunting problems, 356–359
variable traffic density, 232–237

Universe
definition of, 267

Warehouse Problem, 227–232
Weldon's Dice Data, 277–280, 302–310

- 519.2 F947p
Fry, Thornton C
Probability and its
engineering uses,

DEC 17 1952

WHEATON COLLEGE LIBRARY
WHEATON, ILLINOIS

DEMCO

752
cst
4R.